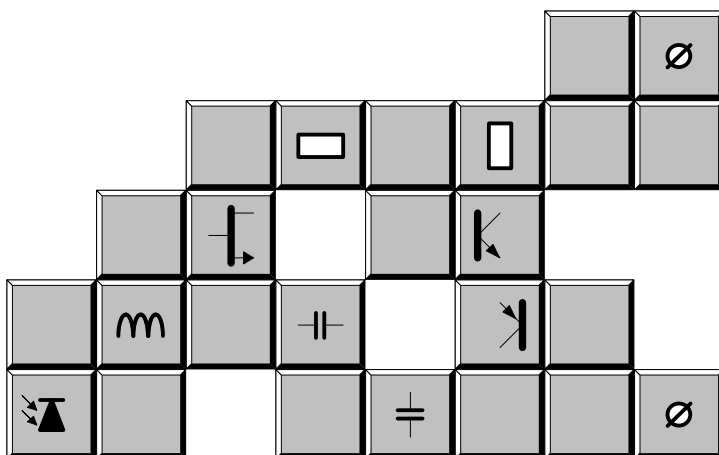

Rob van den Brink

Low-noise Wideband Feedback Amplifiers

an integrated approach to
characterization and design using
microwave and lightwave techniques



Low-noise Wideband Feedback Amplifiers

an integrated approach to characterization and
design using microwave and lightwave
techniques

ruisarme, breedbandig tegengekoppelde versterkers
een geïntegreerde aanpak van karakteriseren en ontwerpen
op basis van microgolf en optische technieken

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft
op gezag van de Rector Magnificus Prof. ir. K.F. Wakker
in het openbaar te verdedigen ten overstaan van een commissie,
door het College van Dekanen aangewezen,
op dinsdag 22 november 1994 te 16.00 uur.

door

Robertus Franciscus Maria van den BRINK

elektrotechnisch ingenieur,
geboren te Oegstgeest

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. ir. R.G.F. Baets

Als toegevoegd promotor treedt op: :
J.L. Tauritz, M.Sc.

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Brink, Robertus Franciscus Maria van den

Low-noise wideband feedback amplifiers: an integrated approach to characterization and design using microwave and lightwave techniques / Robertus Franciscus Maria van den Brink - Leidschendam: PTT Research.

Thesis Technische Universiteit Delft. - With ref. - With summary in Dutch.

ISBN 90-72125-48-7 bound

Subject headings: analog electronic systems / amplifiers.

© 1994 by R.F.M. van den Brink

Subject to the exceptions provided for by law, no parts of this publication may be reproduced and/or published in print, by photocopying, on microfilm, in electronic form, or any other way without the written consent of the copyright owner. The same applies to whole or partial adaptations. The copyright owner retains the sole right to collect from third parties fees payable in respect of copying and/or to take legal or other action for this purpose.

copies of this thesis can be ordered from the secretary of:

PTT Research
P.O. Box 421
2260 AK Leidschendam
The Netherlands

Contents

CONTENTS	0-1
Summary in English.....	0-6
Summary in Dutch	0-8
 INTRODUCTION	 1-1
1.1. Context of this thesis	1-1
1.2. State of the art.....	1-2
1.2.1. Evolution of various fields of study	1-2
1.2.2. State-of-the-art methods for wideband characterization.....	1-4
1.2.3. State-of-the-art methods for wideband feedback design	1-6
1.3. Essentials of this thesis: <i>back to basics</i>	1-9
1.3.1. Characterization	1-9
1.3.2. Feedback design.....	1-10
1.3.3. Overall structure.....	1-11
1.4. Spin-off of this thesis.....	1-12
1.4.1. High performance lightwave receivers.....	1-12
1.4.2. SABEL-CAE design environment.....	1-13
1.4.3. Microwave concepts applied to acoustic applications.....	1-14
1.5. Publications and patents resulting from this work	1-15
 FORMAL DESCRIPTION OF TRANSFER.....	 2-1
2.1. Signal flow observed relative to reference planes.....	2-2
2.1.1. Signal definition in paired conductor links.....	2-4
2.1.2. Equivalent load models.....	2-6
2.1.3. Equivalent source models.....	2-7
2.1.4. Interconnection transformation of waves	2-8
2.1.5. Conclusions.....	2-9
2.2. Blackbox representation of multi-port networks.....	2-10
2.2.1. Multi-port matrix parameters	2-11
2.2.2. Multi-port matrix properties.....	2-13
2.2.3. Multi-port matrix reduction algorithm	2-15
2.2.4. Application of multi-port transfer matrices in circuit simulators.....	2-16
2.2.5. Conclusions	2-17
2.3. Two-port transfer parameters.....	2-18
2.3.1. Two-port matrix parameters.....	2-18
2.3.2. Two-port virtual circuit parameters.....	2-20
2.3.3. Conclusions	2-24

EXTRACTION OF TRANSFER PARAMETERS.....	3-1
3.1. One-port model extraction methods for linear synthesis.....	3-2
3.1.1. Elementary extraction methods of analytical transfer functions	3-2
3.1.2. Extraction methods for equivalent circuit elements	3-4
3.1.3. Extraction of adequate photo-diode impedance-models.....	3-5
3.1.4. Conclusions	3-8
3.2. Two-port extraction methods for transistor models	3-10
3.2.1. Virtual circuit parameters as device representation method.....	3-11
3.2.2. Extraction of transistor parameters using Taylor series expansion.....	3-14
3.2.3. Extraction of parameters with delay using Taylor series expansion.....	3-16
3.2.4. Conclusions	3-17
3.3. Discussion on commonly used transistor models.....	3-18
3.3.1. Discussion on base resistance versus emitter inductance	3-18
3.3.2. Discussion on current transport function vs. diffusion capacitance.....	3-19
3.3.3. Conclusions	3-22
FORMAL DESCRIPTION OF FEEDBACK	4-1
4.1. Superposition analysis of feedback amplifiers.....	4-3
4.1.1. Superposition model parameters	4-3
4.1.2. Superposition flow parameters	4-7
4.1.3. Example of the extraction of superposition parameters.....	4-9
4.1.4. Conclusions	4-10
4.2. Superposition parameter calculation.....	4-11
4.2.1. Calculation of loop gain, forward gain and feedback factor.....	4-11
4.2.2. Calculation of asymptotic and virtual gain and forward leakage.....	4-14
4.2.3. Conclusions	4-15
4.3. Loop gain estimation using single-cut methods	4-16
4.3.1. One-cut loop gain estimation using two-port parameters	4-17
4.3.2. One-cut loop gain estimation using open voltage and current gain	4-20
4.3.3. Conclusions	4-22
4.4. Loop gain representation by poles and zeros	4-23
4.4.1. The purpose of pole-zero patterns.....	4-23
4.4.2. Strategy of pole-zero extraction methods	4-23
4.4.3. Practical demonstration of the pole-zero extractor.....	4-25
4.4.4. Interpretation of pole-zero patterns	4-27
4.4.5. Conclusions	4-28
4.5. Loop gain deflation algorithms.....	4-29
4.5.1. Manual deflation using pseudo delay	4-29
4.5.2. Automated deflation by pole-zero cancellation.....	4-31
4.5.3. Automated deflation using dominant singularities	4-32
4.5.4. Automated overall deflation.....	4-34
4.5.5. Conclusions	4-34
WIDEBAND FEEDBACK SYNTHESIS.....	5-1
5.1. Aperture analysis of feedback amplifiers.....	5-3
5.1.1. Definition of virtual-aperture and effective-aperture.....	5-3
5.1.2. Bandwidth analysis	5-6
5.1.3. Bandpass analysis.....	5-9

5.1.4. Conclusions	5-12
5.2. Compensation techniques for feedback amplifiers	5-13
5.2.1. Profiled compensation techniques in the forward amplifier	5-14
5.2.2. Profiled compensation techniques with intertwined solutions.....	5-16
5.2.3. Phantom compensation techniques in the feedback network.....	5-18
5.2.4. Conclusions	5-21
5.3. Compensation synthesis for feedback amplifiers	5-22
5.3.1. Bandpass synthesis.....	5-22
5.3.2. Compensation algorithm	5-25
5.3.3. Examples of loops with full compensation.....	5-29
5.3.4. Examples of loops with dominant compensation	5-33
5.3.5. Conclusions	5-35

FEEDBACK DESIGN EXAMPLE: LIGHTWAVE RECEIVERS 6-1

6.1. Lightwave receiver configurations.....	6-2
6.1.1. Conventional receiver front-ends	6-3
6.1.2. Improved receivers with current-current feedback.....	6-4
6.1.3. Implementation of receivers using current-current feedback	6-6
6.2. Iterative synthesis focused on maximal bandwidth.....	6-8
6.2.1. Quantifying parasitic effects using circuit stripping techniques	6-9
6.2.2. Cumulative quantification of parasitics using pole-zero techniques.....	6-12
6.2.3. Bandwidth performance of current-current feedback receivers.....	6-13
6.3. Iterative synthesis focused on minimal noise.....	6-14
6.3.1. Noise optimization in the absence of tuning.....	6-15
6.3.2. Noise tuning in absence of overall feedback	6-16
6.3.3. Noise tuning in combination with overall feedback	6-20
6.3.4. Noise performance of current-current feedback receivers.....	6-21
6.3.5. The role of noise measurements in an iterative synthesis.....	6-22
6.4. Conclusions	6-24

FORMAL DESCRIPTION OF NOISE 7-1

7.1. Definitions of signal and noise spectra	7-3
7.1.1. Spectral identification of random signals	7-4
7.1.2. Spectral correlation between random signals	7-6
7.1.3. Analytical noise analysis using complex noise spectra	7-7
7.2. Blackbox representation of noisy multi-port networks	7-8
7.2.1. Description of multi-port circuits including deterministic sources.....	7-8
7.2.2. Description of noisy multi-ports using correlation matrices.....	7-10
7.2.3. Reduction of correlation matrix dimension of noisy multi-ports.....	7-12
7.2.4. Application of noise correlation matrices in circuit simulators.	7-13
7.2.5. Generalized thermal noise theorem for multi-port networks	7-14
7.2.6. Conclusions	7-15
7.3. Two-port noise parameters	7-17
7.3.1. Matrix noise parameters, dedicated to two-ports.....	7-17
7.3.2. Conventional spot noise parameters for two-ports	7-21
7.3.3. Autonomous noise parameters for two-ports.....	7-24
7.3.4. Conclusions	7-26

LIGHTWAVE BASED ELECTRICAL NOISE MEASUREMENTS	8-1
8.1. The art of measuring noise.....	8-3
8.1.1. Pitfalls while measuring noise.....	8-3
8.1.2. Accuracy limits when measuring noise	8-4
8.1.3. Basic definitions.....	8-5
8.2. Multi-level noise source with lightwave noise-tee.....	8-6
8.2.1. Circuit diagram of a noise-tee	8-6
8.2.2. Variable noise source based on a matched noise-tee configuration	8-7
8.2.3. Output noise level of a matched noise-tee configuration.....	8-8
8.2.4. Output impedance variation of a noise-tee configuration.....	8-9
8.2.5. Conclusions	8-11
8.3. Calibration of synthetic noise	8-12
8.3.1. Definition of (spectral) noise-current ratio for synthetic noise.....	8-12
8.3.2. Calibration of synthetic noise with calibrated noise sources	8-13
8.3.3. Calibration of synthetic noise, with shot-noise.....	8-15
8.3.4. Transformation of calibrated noise to arbitrary reference planes	8-21
8.3.5. Mathematical halving of the noise-tee.....	8-23
8.3.6. Conclusions	8-26
8.4. Noise level measurements using matched sources.....	8-27
8.4.1. Basic principles and definitions	8-27
8.4.2. Extraction of input noise level using hot/cold noise sources.....	8-29
8.4.3. Extraction of input noise level using multi-level noise sources.....	8-29
8.4.4. Measurement of lightwave receiver noise	8-32
8.4.5. Alternative lightwave receiver noise measurements.....	8-36
8.4.6. Conclusions	8-38
8.5. Noise parameter measurements	8-39
8.5.1. State-of-the-art measurement methods	8-39
8.5.2. Proposed improvements to state-of-the-art noise measurements.....	8-43
8.5.3. Input noise measurements for mismatched source admittances.....	8-45
8.5.4. Extraction of device noise-parameters	8-47
8.5.5. Extraction of input noise level for arbitrary source admittances	8-49
8.5.6. Conclusions	8-50
LIGHTWAVE SYNTHETIC NOISE GENERATION	9-1
9.1. Basic principles of synthetic noise generation	9-4
9.1.1. Delayed self-homodyne principle with FM modulation	9-4
9.1.2. Example of a practical synthetic noise generator	9-7
9.1.3. Conclusions	9-10
9.2. Practical analysis of synthetic noise generation.....	9-11
9.2.1. Analysis of the spectral envelope of the synthetic noise	9-11
9.2.2. Spectral ripple reduction by noise injection	9-14
9.2.3. Power analysis of synthetic noise.....	9-20
9.2.4. Statistical analysis of synthetic noise	9-22
9.2.5. Conclusions	9-25
9.3. Theoretical analysis of synthetic noise generation.....	9-26
9.3.1. Time domain analysis of the synthetic noise signal.....	9-26
9.3.2. Incoherence analysis	9-29
9.3.3. Spectral analysis far above incoherence threshold	9-34

9.3.4. Statistical analysis far above incoherence threshold.....9-38

9.3.5. Conclusions 9-39

APPENDIX@-1

A. Transformation rules for two-port parameters@-1

B. Algorithms for overdetermined matrix divisions@-2

C. Algorithm for polynomial curve fits.....@-5

D. Algorithm for rational curve fits@-6

E. Algorithm for rational magnitude fit@-8

F. Algorithm for rational delay fit@-9

G. Algorithm for weighed polynomial division@-10

H. Algorithm for dominant deflation of transfer order.....@-11

I. Bessel, Butterworth and Chebyshev transfer functions@-13

J. Algorithm for feedback compensation synthesis.....@-16

K. Definitions of spectra and integral transformations.....@-18

L. Algorithm for extraction of equivalent input noise@-22

M. Noise parameter extraction using hot and cold sources@-24

N. Noise parameter extraction using paired hot and cold sources@-27

O. Acknowledgments.....@-29

P. Bibliography of the author@-30

Q. References.....@-31

R. Index@-37

Summary

Summary in English

The design of wideband amplifiers with ultimate performance requires a radical change of approach. The cumulation of all minor imperfections of individual devices often causes a dramatic reduction of the overall performance. These imperfections (parasitics) are taken into account from the very beginning; so definitely not marginally as it is commonly done.

This thesis introduces innovative methods and tools to facilitate the application of *higher-order* feedback, over *several* amplifying stages, in the presence of deleterious parasitics. Much attention is paid on (1) formal descriptions of transfer, feedback and noise, (2) characterization methods of devices and sub-circuits and (3) design methods for compensation networks in feedback loops.

Formal description of transfer, feedback and noise

The formal descriptions provide unambiguous definitions of (a) multi-port y -, z - and s -parameters, (b) superposition parameters of feedback loops, such as loopgain and aperture, (c) noise spectra and (d) noise parameters. Signal representations, in terms of voltages, currents or waves, are discussed simultaneously.

The concept of virtual circuits is formalized for transfer as well as noise to simplify parameter extraction of device models. Algorithms are described for analyzing transfer and noise properties of arbitrary linear networks. We illustrate, for instance, how to calculate the thermal noise levels in *multi-port* networks from transfer measurements.

Characterization of transfer

Characterization enables the quantification of parasitic effects in terms of (a) multiport transfer parameters, (b) equivalent circuit models, and (c) poles and zeros. This thesis describes methods for measuring these device parameters between well-defined reference planes. A structured method is discussed for tracing relevant parasitics using these measurements (circuit stripping).

Algorithms have been developed for (a) extracting poles and zeros from transfer functions, and (b) reconstructing equivalent circuit models from full one- or two-port measurements.

These algorithms demonstrate that adequate (linear) modeling of transistors does not necessarily result in exotic and complicated transistor models. This approach has, for instance, resulted in a new (small-signal) bipolar transistor model, that is more appropriate to wideband analysis and synthesis than the well-known hybrid- Π model.

Characterization of noise

The development of a new instrument, a lightwave *synthetic noise generator*, has facilitated the application of new methods of noise characterization. Our source generates white noise using *lightwave* principles, and has many advantages with respect

to conventional 50 Ω noise sources. The noise bandwidth of this source can be varied simply, exceeding 10 GHz in bandwidth (hundreds of GHz are feasible).

Novel instruments, principles and algorithms have been developed for (a) calibration of noise sources, (b) noise measurement of electrical and *optical* receivers, and (c) measurement of transistor noise parameters. The feasibility of our measurement methods is demonstrated in practice.

Feedback design

Parasitic effects degrade the dominant behavior of feedback loops. Characterization supplies the parameters for analyzing this degradation, e.g. in a circuit simulator. Adequate compensation networks can minimize this degradation. There is, however, in general no all-embracing answer to the question of how to design these compensation networks, particularly not in combination with harmful parasitics.

This thesis study resulted in robust algorithms for predicting the frequency response of these compensation networks, in terms of poles and zeros. These algorithms are suitable for implementation in (future) circuit simulators. Realistic examples up to 3 GHz demonstrate the effectiveness of our approach.

An integrated approach

All of these methods and tools have contributed to a closer integration of the fields of analog electronics and microwave techniques. The need for an integrated approach to characterization and design finds its origin in the fact that well-known suppression methods of parasitic effects become ineffective in the case of wide bandwidths. It required an in-depth *return to the basics* to combine various methods.

We have carried out an in-depth study of wideband lightwave receivers, resulting in various low-noise receivers using discrete 'low-cost' components. Overall feedback loops spanning two and even three amplifier stages, have been realized with bandwidths exceeding 1 GHz. The applicability of the methods proposed in this thesis have hereby been proven in practice.

(328 pages)

Summary in Dutch

Het ontwerpen van breedbandige versterkers tot op de rand van hetgeen fysisch haalbaar is vereist een radicale verandering van aanpak. De optelsom van alle kleine onvolkomenheden in individuele componenten veroorzaakt vaak een dramatische reductie van de totale kwaliteit. Deze onvolkomenheden (parasiteiten) dienen vanaf het eerste begin volledig in rekening gebracht te worden; dus zeker niet marginaal zoals meestal gedaan wordt.

Dit proefschrift introduceert innovatieve methoden en gereedschappen om *hogere-orde* tekenkoppeling in praktijk te brengen, dus over meerdere versterker trappen tegelijk, onder invloed van hinderlijke parasieten. Veel aandacht is besteed aan (1) formele beschrijvingen van overdracht, tekenkoppeling en ruis, (2) karakterisatie methoden van componenten en sub-circuits, en (3) ontwerp methoden voor compensatie netwerken in tekenkoppelingen.

Formele beschrijving van overdracht, tekenkoppeling en ruis

De formele beschrijvingen leveren ondubbelzinnige definities op van (a) multipoort y-, z- en s-parameters, (b) superpositie parameters van tekenkoppelingen, waaronder lusversterking en apertuur, (c) ruispectra en (d) ruisparameters. Signaal representaties, in termen van spanningen, stromen en golven zijn simultaan besproken.

Het concept van virtuele circuits is geformaliseerd voor zowel overdracht als ruis om parameter extractie van transistormodellen te vereenvoudigen. Algoritmen zijn ontwikkeld voor het analyseren van overdracht- en ruis eigenschappen van willekeurige lineaire netwerken. Zo laten we bijvoorbeeld zien hoe thermische ruisniveaus in multipoort netwerken berekend kunnen worden uit overdrachtsmetingen.

Karakterisatie van overdracht

Karakterisatie maakt het mogelijk parasitaire effecten te kwantificeren in termen van (a) multipoort overdrachts parameters, (b) equivalente circuit modellen, en (c) polen en nulpunten. Dit proefschrift beschrijft methoden voor het meten van deze parameters tussen goed gedefinieerde referentievlakken. Een gestructureerde methode is besproken voor het opsporen van relevante parasieten op basis van deze meetmethoden (circuit stripping).

Er zijn algoritmen ontwikkeld voor (a) het extraheren van polen en nulpunten uit overdrachtsfuncties, en (b) voor het reconstrueren van equivalente circuit modellen uit volledige een- of tweepoort metingen. Deze algoritmen laten zien dat een goede (lineaire) modelering van transistoren niet noodzakelijk resulteert in exotische en gecompliceerde transistor modellen. Deze aanpak heeft bijvoorbeeld een nieuw (klein-sigitaal) bipolair transistor model opgeleverd, dat veel beter geschikt is voor breedbandige analyse en synthese dan het bekende hybride-II model.

Karakterisatie van ruis

De ontwikkeling van een nieuw instrument, een *synthetische ruisgenerator*, heeft de toepassing van nieuwe ruiskarakterisatie methoden mogelijk gemaakt. Onze bron genereert witte ruis op basis van *optische* principes, en heeft vele voordelen ten opzichte van conventionele 50 Ω ruisbronnen. Zo is de ruisbandbreedte van deze bron eenvoudig te variëren, tot meer dan 10 GHz bandbreedte (vele honderden GHz zijn mogelijk).

Nieuwe instrumenten, principes en algoritmen zijn ontwikkeld voor (a) het calibreren van ruisbronnen, (b) ruismetingen aan elektrische en *optische* ontvangers, en (c) het meten van transistor ruisparameters. De toepassing van onze meetmethoden is in de praktijk getoetst.

Ontwerp van tegenkoppelingen

Parasitaire effecten degraderen het dominante gedrag van tegenkoppelingen. Karakterisatie levert de parameters voor het analyseren van deze degradatie, bijvoorbeeld met een circuitsimulator. Goede compensatie netwerken kunnen deze degradatie minimaliseren. Er bestaat echter geen alomvattend antwoord op de vraag hoe deze compensatie netwerken ontworpen moeten worden, met name niet in combinatie met hinderlijke parasieten.

Dit promotieonderzoek heeft robuuste algoritmen voortgebracht voor het voorspellen van de frequentiekaracteristiek van compensatie netwerken, in termen van polen en nulpunten. Deze algoritmen zijn geschikt voor implementatie in (toekomstige) circuitsimulators. Realistische voorbeelden tot 3 GHz tonen de kracht van onze aanpak aan.

Een geïntegreerde aanpak

Al deze methoden en gereedschappen hebben de vakgebieden van analoge elektronica en microgolfttechniek dichter bij elkaar gebracht. De noodzaak van een geïntegreerde aanpak van karakteriseren en ontwerpen komt voort uit het feit dat bekende onderdrukkings principes van parasitaire effecten tekort schieten bij het ontwerpen van zeer breedbandige schakelingen. Het vereiste een grondige *terug-naar-het-begin* aanpak om verschillende methoden te kunnen combineren.

We hebben veel onderzoek verricht naar breedbandige optische ontvangers, hetgeen geresulteerd heeft in verschillende ontvangers op basis van goedkope discrete componenten. Algehele tegenkoppeling, over twee tot drie versterkertrappen tegelijk, is gerealiseerd voor bandbreedtes tot voorbij 1 GHz. De methoden die ontwikkeld zijn in het kader van dit proefschrift hebben zich hiermee in de praktijk bewezen.

(328 pagina's)

Chapter 1

Introduction

1.1. Context of this thesis

Fiber optic telecommunication systems have been shown to be suitable for the transport of data rates exceeding 10 Gb/s (gigabits per second) over a single fiber. These bit rates are feasible when transported in several parallel streams through multichannel systems¹, or truly sequentially through single channel systems² using ultra fast opto-electronic components.

The relevance of particular system performance specifications, such as bandwidth, sensitivity and distortion, is of course application dependent.

- Bandwidth is a measure for transport capacity. Bandwidth is applicable in (1) trunk lines (point to point links) for the transport of high volumes of data and in (2) networks where the medium must be shared³ by multiple users, nodes and/or subscribers.
- Sensitivity is a measure for the maximum distance that can be bridged in a trunk line or the maximum number of subscribers that can be simultaneously connected to passive optical networks [103].
- Link distortion leads to intermodulation that limits the performance and the number of TV-channels in analog lightwave video links [106]. In digital links, distortion deteriorates the bit error rate and burst-performance of the system [108].

System performance requirements set high demands on lightwave transmitters and receivers. In subscriber networks, the use of expensive exotic components for meeting performance requirements is viable only on the local exchange side. This is because investments at this side of the network are shared with multiple subscribers. On the other hand, the costs on the subscriber side of the lightwave network are decisive for the economical viability of the network. This justifies the in-depth investigation of low-cost lightwave receiver and transmitter realizations, while optimizing performance [101].

Improved performance, realized using low-cost technology, increases the economical feasibility of lightwave subscriber networks. This can be illustrated as follows.

- Similar performance, realized with cheaper components, reduces overall system cost.
- Increasing sensitivity while maintaining system cost increases the number of subscribers that can be simultaneously connected to passive optical networks, This better shares system investments.
- Increasing bandwidth while maintaining equal cost reduces the number of lightwave channels (fibers or wavelengths) needed to fulfill system transport capacity

¹ As in wavelength division multiplex systems (WDM) or in coherent systems [105,107].

² As in direct detection systems [102].

³ Electrical and optical ATM networks (asynchronous transfer mode) are examples of networks sharing transport capacity [104].

requirements. This increases access to sophisticated services for the majority of users.

The tools and methods described in this manuscript facilitate the design of wideband amplifiers with optimal performance. Although the examples are focused on lightwave receiver design, and the design examples are realized using discrete components mounted on low-cost glass-epoxy prints, the methods and tools are more widely applicable.

1.2. State of the art

The present trend in lightwave transmission systems to increasingly broader bandwidths places unique demands on the design and realization of the associated electronics. The symbiotic melding of techniques, drawn from traditional analog electronics as well as microwave circuits, forms the primary terrain of this study. It resulted in a new field we have chosen to call *wideband⁴ electronics*.

A seamless integration of analog and microwave electronics is not as simple as one might expect since both disciplines use different approaches. In practice, microwave engineers and analog electronic engineers speak different 'languages'. Recent developments, however, in automated (vector) network analyzers and microwave circuit simulators have significantly simplified this integration, stimulated by the present trend in lightwave transmission systems.

This section provides a short overview of these fields and discusses state-of-the-art aspects of wideband characterization and feedback design.

1.2.1. Evolution of various fields of study

The problems attendant to the development of large scale telephone links in the 1920's stimulated the systematic study of electronic systems. Voltage, current, and wave concepts, as well as circuit theory, transmission line theory and feedback theory, are all known aspects of electronic design. The complexity of electronic application led to the rise of many different fields of study, focusing on specialized topics. Examples are analog electronics, microwave electronics, circuit simulation, opto-electronics, etc.

Since an exact definition of each field of study does not exist, this section tends to be descriptive of trends.

Analog electronics forms the basis for many aspects of electronic design. This discipline has affinity with the type of analog circuits used in various audio and video applications. Analog engineers are well acquainted with feedback loops at the circuit level.

The in-depth study of feedback amplifiers, first introduced in 1934 by Black [401], has resulted in a variety of basic feedback theories. These theories include the general feedback theory of Bode [403], developments in the early 1950's of root-locus analysis

⁴ The words *wideband* and *broadband* are synonymous.

methods⁵, compensation of loop stability using phantom zeros [404] in 1961, and the structured design (synthesis) of feedback amplifiers proposed by Nordholt [406] in 1983. Many applications of feedback are focused on frequencies below 100 MHz and share more than a decade in bandwidth. The first column of figure 1.1 shows a survey of characteristic aspects, tools and techniques drawn from practice.

analog electronics	microwave electronics
<i>typical aspects</i> many applications below 100 MHz many applications more than a decade bandwidth many application using feedback feedback usually covering several stages voltages and currents preferred Z-, Y-, H-parameters often used sources with zero or infinite impedances lumped ⁶ elements preferred	<i>typical aspects</i> many applications above 1 GHz many applications within an octave bandwidth many applications using resonators feedback usually avoided waves preferred S-, T-parameters often used sources with matched impedances distributed ⁷ elements often used
<i>powerful tools and techniques</i> circuit simulator Spice/Pspice/... structured synthesis of feedback amplifiers pole-zero patterns root-locus analysis	<i>powerful tools and techniques</i> Touchstone/Libra/MDS/SuperCompact network analyzer stability circles in Smith charts measurements relative to reference planes full two-port transfer parameter analysis full two-port noise parameter analysis microstrip and stripline print layout

Fig 1.1 Some typical aspects of analog and microwave electronics. Both disciplines are well-acquainted with ABCD parameters.

Microwave electronics is a specialized discipline, having its origin in the application of extraordinarily high frequencies. A different type of engineering was required to handle these frequencies. The invention of the radar just prior to World-War II was based on the application of resonance techniques and power matching.

Powerful design concepts that are often used in this field are waves related to well-defined reference planes and scattering parameters (s-parameters). The growing interest in microwaves resulted in 1939 in the development of the Smith Chart at Bell Laboratories by Philip Smith. In the years 1942 to 1946 a rather intensive and systematic exploitation of microwave problems was carried out at the Radiation Laboratory of MIT. In the years 1946-1952 these studies led to a series of 28 volumes dealing with microwave and related techniques, including [203,204].

⁵ Although loop gain variation is most widely used in root-locus analysis, there is no good criterion for selecting this parameter for the best available choice. Variation of circuit element values (pole-zero position diagrams) is another possibility.

⁶ A *lumped* element is an element, so small that wave propagation effects may be ignored. It is modeled as if it has zero 'volume'.

⁷ A *distributed* element is an element in which wave propagation effects cannot be ignored. An example is a transmission line. The volume of distributed elements cannot be ignored with respect to the wavelength of interest.

The increasing application of s-parameters prompted Deschamps [212] for development of a graphical method for the extraction of s-parameters from reflection measurement data. In 1965, Kurokawa [210,211] published a generalized s-parameter theory and two years later the first automatic network analyzer [213,214] was introduced for directly measuring two-port s-parameters. This seminal design remained in production for about twenty years. In the 1980's the popularity of s-parameters increased because of the appearance of new network analyzers which performed s-parameter measurements with ease.

Many microwave applications are focused on frequencies above 1 GHz, and are concentrated in octave bandwidths. Although *absolute bandwidths* may exceed several GHz, *relative bandwidths* in most applications are quite small.

The second column of figure 1.1 surveys microwave aspects, tools and techniques drawn from practice.

Circuit simulators have been developed for specific analog electronic applications as well as specific microwave applications. Since the mid-1970's a number of computer programs for circuit simulation have become commercially available. The SPICE family [127] with many members [407] is probably the most well-known family of simulators. This family supports linear analysis as well as non-linear analysis using time-domain analysis. The analysis is based on equivalent circuit models of semiconductor devices [303]. Recent developments in symbolic circuit simulators enable automated design of analog integrated circuits [511,513].

Microwave applications require another approach. In general, models tend to fail when the frequency of interest increases, due to unknown parasitic effects. The development of simulators combining device models with device measurements (s-parameters in tabular form) resulted in typical microwave simulators. Examples are Super-Compact/Harmonica, Touchstone/Libra [124,125] and MDS [126].

These commercially available microwave simulators are more powerful than SPICE in that (1) they give full-support to the application of s-parameters in tabular form and distributed elements, and that (2) they facilitate iterative optimization of circuit parameters, guided by user-defined design goals. Harmonic-balance algorithms are used for non-linear steady-state analysis to reduce the overall calculation effort (compared to time domain analysis).

Microwave simulators are equipped with extensive libraries of models of microstrip and stripline elements. These models facilitate the simulation of planar microwave structures. More recently, the development of two- and three-dimensional EM-field solvers has significantly broadened the scope of these in essence physical simulations.

1.2.2. State-of-the-art methods for wideband characterization

Over the years, the increasing transition frequency of semiconductor devices has shifted amplifier and lightwave receiver⁸ bandwidths beyond 1 GHz. Designing low-noise wideband circuits, for use at high frequencies, requires reliable knowledge of parasitic effects. This holds for transfer as well as noise aspects.

⁸ State-of-the-art examples are lightwave receivers, such as the commercially available HP70810b, ranging from 100kHz–22GHz.

Desired effects as well as parasitics are quantified using techniques which we have chosen to call *characterization*. Characterization starts with measurements on devices or subcircuits, relative to well-defined reference planes (see section 2.1). The required information is subsequently reconstructed from the measurement data by transforming it into a most appropriate form, followed by extraction of the parameters of interest. Examples of these parameters are (1) element values of equivalent circuit models and (2) poles and zeros of equivalent transfer functions (see chapter 3).

This section 1.2.2 considers in short parasitics, high frequencies, and state-of-the-art characterization methods.

What are parasitic effects?

An exact definition of parasitic effects does not exist in our case. It is a collective term for all frequency-dependent effects in a circuit that are undesired by the designer. As a rule of the thumb, all equivalent circuit elements that change a transfer function of interest are parasitic elements when the change is undesired by the designer. Deleterious parasitics are parasitics for which the change in transfer exceeds predefined quality limits.

If parasitic effects are harmful or not, is application dependent. This 'definition' is obvious when parasitics prevent feedback amplifiers from stable operation. In the case that parasitics introduce additional ripple in the overall transfer, the quality requirements of the application determine whether these parasitic effects are harmful or not.

What are high frequencies?

A similar discussion holds for terms such as *low* frequencies, *high* frequencies, *narrowband* and *wideband*. In this study, we relate these concepts to deleterious parasitic effects because most parasitic effects increase with frequency. The various impedance⁹ values, the mechanical construction as well as the complexity of a circuit are decisive whether high-frequency aspects play a role or not. For example, a poorly designed print layout may suffer from deleterious parasitics above 1 MHz, while proper redesign may shift this limit to above 100 MHz. In the first example, 10 MHz would then be a 'high' frequency and in the second example a 'low' frequency¹⁰.

In this study, the word *wideband* is reserved for frequency ranges involving high frequencies, that range over more than a *decade*. The word *narrowband* is reserved¹¹ for frequency ranges less than an *octave*.

State-of-the-art transfer characterization

Two-port transfer measurements, relative to well-defined reference planes, have become relatively simple since modern intelligent microwave network analyzers (mid 1980's) and lightwave component analyzers (early 1990's) have become commercially available

⁹ For instance, a current to voltage conversion using a resistor R is considered to be a low-frequency application for frequencies $\omega \ll 1/(RC)$, in which C denotes a parasitic shunt capacitance. It means that when $C=1$ pF, the conversion is a high-frequency one above 1 MHz when $R=100$ k Ω . On the other hand, the application is a low-frequency one up to 1 GHz when $R=100$ Ω .

¹⁰ Some microwave engineers refer in waveguide applications to all signals below 1 GHz as "dc", since it is simple to transport them via wires. This illustrates how relative the notions "low" and "high" frequency are.

¹¹ In classical microwave applications, bandwidth of more than say 20% of the center frequency are usually referred to as *wideband*. This illustrates how relative the notions "narrowband" and "wideband" are.

and. They cover a wide range of frequencies. Network analyzers usually measure two-port parameters in terms of s-parameters.

Currently, many manufacturers are specifying their semiconductor devices on the basis of s-parameter measurements. An increasing number of manufacturers provide SPICE parameter sets as well. They represent the parameters of equivalent transistor models, applicable in SPICE family of circuits simulators.

A variety of similar transistor models for circuit *analysis* are described in the literature. In general, these models are inconvenient for circuit *synthesis*¹². The complexity of synthesis forces the use of models that are *as simple as possible*. It requires models with a minimum number of parameters, yet still adequate over the full frequency band of interest.

Prior to this study, a structured method for extracting these synthesis models was lacking. Our techniques for transforming measured s-parameters into more dedicated formats (see chapter 2, virtual circuit parameters) and pole-zero techniques (see chapter 3) have proven to be successful in practice for this purpose.

This work introduces general methods for extracting a minimum number of poles, zeros and model parameters from device measurement.

State-of-the-art noise characterization

The ease of transfer measurements, relative to well-defined reference planes, does not hold for similar two-port noise measurements.

Currently, instruments for measuring (electrical) input noise level under matched impedance conditions are commercially available covering a wide range of frequencies. Some setups for full two-port (electrical) noise measurements are commercially available, however they focus on the microwave frequency band. Prior to this study, instruments for reliable optical noise measurements on lightwave receivers were lacking. For this reason, our study of noise characterization was mainly focused on the measurement of noise. Noise modeling and low noise design aspects are not discussed in great detail.

This work introduces an innovative lightwave noise source and various calibration and measurement methods for input noise measurements on lightwave receivers and two-port noise measurements on electrical amplifiers.

1.2.3. State-of-the-art methods for wideband feedback design

Feedback loops facilitate the stabilization of gain in amplifiers over several frequency decades. The greater the bandwidth, or the number of cascaded amplifier stages covered by a single loop, the more effective this stabilization may be. On the other hand, deleterious parasitic effects prevent the circuit from proper operation, especially when designing higher order feedback loops for high frequencies. These obstacles make wideband feedback design complicated and a specialist discipline.

This section 1.2.3 considers higher order feedback loops, and state-of-the-art feedback design methods.

¹² The purpose of analysis is to find the properties of an extant circuit. Circuit simulations as well as measurements are adequate methods. The purpose of synthesis is to find a circuit that meets a set of pre-defined conditions. This is the reverse problem.

What are higher order feedback loops?

Feedback enables the stabilization of gain by using passive linear components, to improve the overall transfer function. The more amplifying stages are covered by an overall feedback loop, the more effective this stabilization will be.

The order of a feedback loop equals the dominant transfer order of the loop gain. The more amplifying stages in the loop, the higher the feedback order is, and the more complicated the design will be to ensure stable operation of the loop. Therefore the loop order in wideband circuits is in practice kept as low as possible, usually covering a single amplifying stage. To simplify matters, feedback loops can be divided into first-order and higher-order loops.

- *First-order feedback loops* are simple in nature and usually cover one stage (local feedback). A stage can be a single transistor or a transistor decoupled by a common-base (gate) configuration¹³. This approach minimizes the transfer order of the loop gain and avoids the need for compensation. Therefore, ultra wideband first-order loops have been reported for lightwave receivers exceeding 10 GHz in bandwidth. An example is a 14 GHz (monolithic integrated) lightwave receiver [815], using local transimpedance feedback.
- *Higher-order feedback loops* require adequate compensation networks to prevent oscillatory behavior. They are fed back covering more than one amplifier stage (overall feedback), in which the gain of each individual stage is maximized by avoiding local feedback. The application of these loops is usually restricted to frequencies below 100 MHz. The wider the bandwidth of interest, the more the overall stability must rely on adequate compensation and the more sophisticated the loop will be.

Using higher-order feedback loops is advantageous because well-implemented loops improve the flatness and the linearity of the overall transfer function.

State-of-the-art feedback design

The use of higher order feedback loops places high demands on the circuit design. Therefore, many-ultra wideband receivers use no feedback at all, and rely on an established microwave approach based on carefully designed resonators and matching networks.

The use of higher-order feedback loops is a typical analog electronic design gambit. Nordholt [406], inspired by the work of Cherry and Hooper [405], developed a systematic and straightforward synthesis method to design amplifiers using higher-order feedback loops. It is a hierarchical design method, in which the overall circuit is developed in a series of design steps.

One aspect of this synthesis is that it suppresses deleterious parasitic effects introduced by additional circuitry, such as cascoded¹⁴ amplifier stages or long-tailed¹⁵ transistor pairs (active de-coupling). The advantage of this technique is that the designer may use

¹³ In general, a cascode stage neutralizes the parasitic 'Miller' capacitance, compared with a single transistor stage. It has inverted gain. A long-tailed pair is a transistor pair that has similar performance and non-inverted gain.

¹⁴ A cascode configuration is a cascade of a CE - CB configuration (common Emitter, common Base).

¹⁵ A long tailed pair a cascade of a CC - CB configuration (common Collector, common Base). In differential applications it is also called a balanced transistor pair.

simplified models for the individual circuit devices, to handle the complexity of the overall synthesis.

The above design approach is very effective for relatively low frequencies¹⁶, up to say 1% of the transition frequency f_T of the amplifying devices. Designers, well-skilled in the art, may even apply this approach for significantly higher frequencies, up to perhaps 10% of f_T . The influence of parasitics on the signal flowing through the individual devices, nevertheless, limits the designability of the overall feedback loop. This is because active decoupling tends to fail in wideband feedback loops, for instance, due to secondary parasitics introduced by the active decoupling circuits.

Many higher-order feedback amplifiers have been developed successfully for frequencies below 10 MHz using Nordholt's method. Higher-order feedback realizations exceeding 100 MHz in frequency are much less prevalent. Nevertheless, third order loops covering 500 MHz in bandwidth and second order loops exceeding 1 GHz in bandwidth are applicable, as is demonstrated in chapter 6.

Improved feedback analysis

In general, the complexity of the overall design increases with frequency, so that it is a serious challenge to design *higher-order* feedback loops for multi-GHz amplifiers, in the presence of dominant parasitics. Microwave circuit simulators simplify the analysis of wideband circuits. They give full-support to the application of s-parameters and distributed elements, for analyzing the deterioration of overall performance by parasitic effects. However, full support for pole-zero analysis of loop gain and other superposition parameters (see chapter 4) is lacking in current microwave simulators. Additional analysis tools for the *synthesis* of wideband feedback circuits have been developed in the course of this study.

|| This work introduces various supplementary algorithms for (future) circuit simulators to analyze feedback loops in terms of superposition parameters and root locus plots. ||

Improved feedback synthesis

Nordholt's synthesis provides powerful methods for compensating feedback loops, in the case of instability. When parasitics are dominant and their suppression using active decoupling fails, Nordholt's compensation synthesis may fail.

Prior to this study, the question of dealing with dominant parasitic effects in feedback loops, was an unanswered one. Computer assisted synthesis methods for the design of higher-order wideband feedback loops, in the presence of deleterious parasitics, are dealt with in depth in chapter 5.

|| This work introduces an innovative synthesis algorithm to synthesize the transfer function of compensating networks required for stable feedback loops. ||

¹⁶ When designing amplifiers using transistors with 5 GHz transition frequency, 1% means 50 MHz bandwidth, and 10% means 500 MHz amplifier bandwidth.

1.3. Essentials of this thesis: *back to basics*

The design of wideband amplifiers with ultimate performance required a radical change of approach. The cumulation of all minor imperfections of individual devices often cause a dramatically reduction of the overall performance.

These imperfections (parasitics) had to be taken into account from the very beginning; so definitely not marginally as it is commonly done. This required an in-depth *return to the basics* to enable an *integrated approach* to characterization and design. It resulted in profound formal descriptions of transfer, feedback and noise.

Of particular importance is the emphasis on full *characterization* of devices and subcircuits. Measurements aren't just used for verifying performance when circuit design has completed. They form an integral part of the overall design procedure and are important for quantifying parasitics. These parasitics are identified by extraction from wideband full two-port measurements. This study has led to a variety of characterization methods for transfer as well as noise.

Of equal importance is the computer assisted *design* of feedback loops, taking full account of the parasitics. State-of-the-art design methods are applicable to the synthesis of an initial design from scratch. *Supplementary* to the methods of Nordholt [406], our algorithms provide, in the presence of dominant parasitics, the transfer functions of compensation networks required for stable feedback operation.

1.3.1. Characterization

Back to basics meant in practice that most of this thesis is focused on *characterization*. Transfer measurements are discussed and demonstrated with devices mounted on standard printed circuit boards. For noise characterization (electrical and optical), a promising lightwave noise source has been developed, including innovative test setups and measurement protocols. New methods and algorithms are introduced for parameter extraction from transfer and noise measurements.

This work has dealt with various obstacles to state-of-the-art characterization:

- The use of formal descriptions of transfer, feedback and noise brought to light that often used concepts are ambiguous, mixed up, or restricted to limited conditions. Chapters 2, 4 and 7 provide unambiguous definitions, discuss voltages, currents and waves simultaneously, and describe the relationship between various equivalent parameter representations.
- The transfer measurements used have clearly demonstrated that commonly used hybrid- Π transistor models tend to fail above 10% of the transition frequency. This in turn has led to questions concerning the validity of the diffusion capacity concept in these models. Chapter 3, discusses extraction methods leading to more suitable transistor models. They demonstrate that adequate (linear) modeling of transistors does not necessarily result in exotic and complicated transistor models.
- Our noise measurements have confirmed how tricky it is to carry out meaningful noise measurements. Noise levels may in practice be as much as 5.5 dB too low: 2~2.5 dB when spectral measurements are only corrected for resolution bandwidth

effects and another 3 dB when double sided intensity spectra are measured without further notice. These are serious pitfalls because many publications on noise performance of lightwave receivers do not properly specify *what* noise quantity is measured and *how* its value is reconstructed from measurements. Universally accepted conventions on these topics are lacking.

Chapter 8 introduces innovative measurement methods for noise, based on the lightwave noise source discussed in chapter 9. These methods make feasible the development of simple and yet reliable noise models of semiconductor devices.

1.3.2. Feedback design

The hierarchical design method of Nordholt [406] follows a *top-down* approach. When parasitics are adequately suppressed by active decoupling, idealized block diagrams (superposition model) are adequate for describing transfer and noise performance of amplifiers.

This top-down approach is very effective in the initial design phase, although it may fail when active decoupling fails. These are typical obstacles when designing feedback amplifiers exceeding $f_T/100$ or $f_T/10$ in bandwidth. In general, designers get into trouble when suppression of parasitics becomes ineffective. The established approach is: keep it simple, avoid higher order feedback loops, use superior technology and accept limited performance.

We propose an additional *bottom-up* approach to handle higher order feedback loops when active decoupling fails. The resulting iterative improvement of the design, is illustrated in figure 1.2.

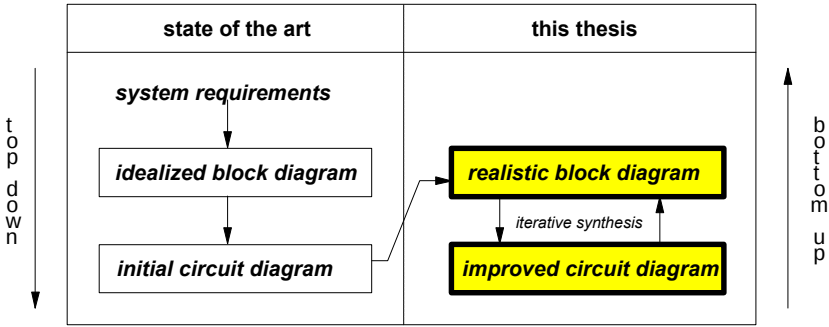


Fig 1.2 A top-down design approach, using parameters of idealized block diagrams, is effective in the initial design phase. In wideband amplifiers, parasitics are harmful and require an additional iterative approach. In this approach, the parameters of realistic block diagrams are extracted from measurements and simulations, and used to improve the design.

This study resulted in computer algorithms for extracting the parameters of a block diagram (superposition model) from first-try designed circuit diagrams. Parasitics are fully taken into account increasing the accuracy of extracted parameters.

These parameters are supplied to another algorithm that analyzes the loop stability and calculates the transfer function (in terms of poles and zeros) of adequate compensation networks. This is valuable information for designers, who can use this to improve their initial design.

Our approach is required when circuit analysis indicates that the overall circuit performance significantly deviates from using ideal block diagrams. Possible conditions include: oscillations, ringing, ripple, smaller bandwidths or higher noise levels.

1.3.3. Overall structure

This work is divided roughly in two parts: chapters 2 to 5 deal with various transfer aspects and chapters 7 to 9 with noise. Chapter 6 considers the application of our approach to the design of lightwave receivers, linking in effect the two parts.

(A) Transfer characterization and feedback design:

- Chapter 2 starts with a formal description of electronic networks in terms of two-ports and multi-ports, and discusses appropriated measurement methods.
- Chapter 3 introduces innovative algorithms for extracting parameter sets of device models from two-port measurements using *poles and zero* techniques. This has led to an improved linear model for the bipolar junction transistor; simple and yet applicable up to the transition frequency f_T .
- Chapter 4 introduces algorithms for extraction of the *superposition parameters* in the presence of parasitic effects. The associated superposition model provides a realistic block diagram for the analysis of feedback amplifiers.

Various algorithms are introduced in chapter 4 for evaluating the loop gain of feedback amplifiers, to represent them by pole-zero descriptions and to simplify this description to the dominant form.

- Chapter 5 discusses compensation methods for neutralizing dominant poles and zeros in feedback loops. This chapter introduces a powerful algorithm for the synthesis of compensation networks, in terms of poles and zeros.
- Chapter 6 illustrates the application of the tools of this study in the support of lightwave receiver design. These receivers have been constructed using second order feedback loops yielding bandwidths in excess of 1 GHz, and using third order feedback loops (500 MHz). This chapter discusses the discrepancy between measured and simulated receiver noise to illustrate the importance of reliable characterization of noisy devices.

(B) Noise characterization:

- Chapter 7 discusses the formal description of noisy devices in terms of two-port and multi-port noise parameters. This chapter relates *matrix noise parameters* (convenient for noise analysis) with *spot noise parameters* (commonly used to specify semiconductor devices) and introduces *autonomous noise parameters* (convenient for evaluating device models). Calculation methods and generalized methods for evaluating thermal noise in passive networks are discussed in detail.
- Chapter 8 introduces new methods for measuring electrical noise using lightwave principles. This chapter demonstrates the superior performance of our *noise-tee* compared to conventional 50Ω noise sources, and discusses methods for calibrating

the associated lightwave noise source. Methods for measuring lightwave receiver noise and measuring two-port noise parameters are demonstrated.

- Chapter 9 introduces the *lightwave synthetic noise source*, a new instrument for generating white electrical noise to support the measurement methods of chapter 8. One of the advantages of this lightwave source is that the noise is concentrated in a user definable bandwidth (ranging from tens of MHz to hundreds of GHz). This prevents overload problems to the device under test.

1.4. Spin-off of this thesis

The pole-zero extraction tools that have been developed in this study are more widely applicable than the field of wideband electronics. The transfer function of any (linearized) physical process can be reconstructed from measured data using these tools. In addition, our compensation algorithms yield poles and zeros that are applicable for controlling the stability of any linear feedback process.

Currently, our study has resulted in (1) high performance lightwave receivers, (2) a design environment, and (3) several acoustic applications.

1.4.1. High performance lightwave receivers

This study resulted in various lightwave receivers using current-current feedback. Their advantages and performance are discussed in chapter 6.

We were the first to propose this feedback principle for wideband receiver design [608], and demonstrated the applicability of this concept for frequencies exceeding 1 GHz. Applications have been found in various lightwave transmission experiments, including the European projects ESPRIT 2054 (UCOL), RACE 2024 (BAF [108]) and PTT-Research projects such as COSNET [107].

Focusing on the ultimate performance of these receivers yielded the understanding that wideband synthesis must be based on sophisticated measurement techniques and powerful computer algorithms. This has resulted in tools and concepts that are more widely applicable than lightwave receiver design.

The application of lightwave receivers using discrete components

All design examples in this work are restricted to discrete components mounted on low-cost glass-epoxy printed circuit boards (PCB's). Integrated circuit solutions or thin film solutions have not been investigated, although our design theories and tools can also be applied to these technologies.

The wide bandwidth results of our receivers illustrate that it is not mandatory to use chip technology when designing wideband high-performance circuits. Nevertheless, significant improvements in performance are to be expected when using thin-film technology in combination with chip technology. This holds in particular when choosing a foundry processes with the highest f_T available.

When bandwidth demands are fulfilled, low-cost aspects may be an argument for using chip technology. We note that it is, however, not so obvious that chip-technology is always pre-eminently suited for low-cost wideband applications. This is best illustrated as follows:

- Discrete SMD¹⁷ components have been optimized for reduced costs when circuits are produced in high volumes. Consumer electronics, in fact, frequently uses discrete components in combination with ASIC's.
- Furthermore, discrete technology enables the combination of widely differing devices, such as GaAs FET's with PNP *and* NPN Si-BJT's. Therefore, a designer can take full advantage of this benefit and can combine the most suitable devices.

It is perhaps indicative that, as we enter the last quarter of 1994, fully integrated low-cost lightwave receivers are still commercially unattractive when compared with receivers using discrete technology. Commercially available monolithic transimpedance amplifiers (>1 GHz) are currently intended for use with photo diodes with relatively low capacitance (~0.3 pF). The price of these diodes ($\lambda = 1300\text{-}1500\text{ nm}$) is often indicative for the price of the overall receiver. Our 1 GHz current-current feedback receivers enable the use of low-cost photo diodes with relatively high capacitance (~1.5 pF).

Moreover, state-of-the-art television tuners are produced in high volumes, and most of them use many discrete components. The above discussion indicates that discrete components and chips are competitive technologies when designing 1 GHz lightwave receivers. Therefore, restricting our use to discrete components is a realistic one, although we do not wish to exclude other technologies.

1.4.2. SABEL-CAE design environment

The algorithms proposed in this work are implemented in software and tested using actual measured data. From the very beginning of this study (1985) we realized the importance of linking measurement instruments with simulation software. Over the years, we realized the SABEL-CAE¹⁸ design environment [111].

A complete discussion of this environment is beyond the scope of this text. Some hardware aspects of this environment are shown in figure 1.3. This short overview is, furthermore, restricted to the description of self-made elements:

- *Copydata* is a Pascal program [111] for data acquisition. It links various instruments with a PC and provides the data in various file formats. It supports file formats, for signal processing with Matlab®[123], for linear circuit analysis with Touchstone® [124] and for non-linear circuit analysis with Libra®[125]. These three commercially available software tools are embedded in the design environment.
- *Locus* is a Pascal program for VAX-VMS to facilitate our study on feedback loops. Feedback analysis is discussed in chapter 4 and feedback synthesis in 5. The software facilitates (1) analytical circuit analysis for calculating poles and zeros, (2) pole-zero extraction from measured data using a mix of Marquardt iteration, and Gauss-Newton iteration, and (3) root-locus analysis for feedback systems. [113,114,115,116,117,118]
- Most algorithms described in this text are implemented as Matlab® applications (a fourth generation mathematical computer language). This has resulted in an extensive collection of general purpose data processing tools: the microwave toolbox *MW-tools*. [112]

¹⁷ SMD = surface mounted devices.

¹⁸ SABEL-CAE = Synthesis and Analysis of Broadband Electronics using Computer Assisted Engineering

- The lightwave *noise testset* is an instrument that has evolved from our study on minimizing noise in lightwave receivers. Noise measurements are discussed in detail in chapter 8 and 9.

For more information on this environment see [101,109,111] and various related internal reports (see [111] for a complete overview).

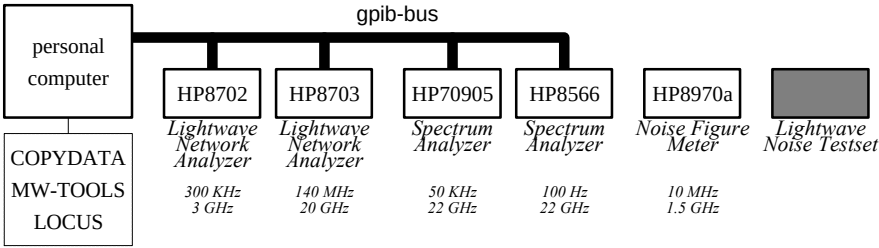


Fig 1.3 Hardware part of the SABEL-CAE design environment.

1.4.3. Microwave concepts applied to acoustic applications

One of the basic concepts in this work is a blackbox description of devices and sub-circuits when adequate equivalent models are not available. This approach is not restricted to electrical application and is also applicable to a variety of other technical disciplines.

For instance, we have introduced two-ports and s-parameter concepts in acoustics [119], which can be regarded as a spin-off of this work. Furthermore, we have developed an *acoustic* network analyzer [120] and applied microwave calibration techniques to it [121,122]. The acoustic equivalent of shorts, opens and loads are respectively (1) pipes with covers, (2) open pipes and (3) pipes with absorbers.

We applied microwave measurement techniques to measure *acoustic* transmission in ducts (pipes). These methods were previously unknown in acoustics. Microwave simulators, such as Touchstone®[124], have proven to be valuable in designing matching networks in acoustical oscillators. Microwave matching techniques and microwave circuit simulators (linear or non-linear) have successfully used for designing thermo-acoustic energy converters. Other possible applications are acoustical filters in air conditioning systems, and exhaust systems for combustion engines.

This illustrates the power of an interdisciplinary design approach.

1.5. Publications and patents resulting from this work

Patent applications

- [803] R.F.M. van den **Brink**: *Method for lightwave noise measurements*. Patent application, Europe: EP-0522614, USA, June 1991.
- [901] R.F.M. van den **Brink**: *Lightwave noise source*. Patent application, The Netherlands: NL-9300347, Feb 1993.

International publications (electronics)

- [608] R.F.M. van den **Brink**: *Optical receiver with third-order capacitive current-current feedback*, Electronics Letters, vol 24 no. 16, pp 1024-1025, Aug 1988.
- [801] R.F.M. van den **Brink**, E.**Drijver**, M.O van **Deventer**: *Novel noise measurement setup with high dynamic range for optical receivers*, Electronics Letters, vol 28 no. 7, pp 629-630, March 1992.
- [802] R.F.M. van den **Brink**: *Novel electrical noise source based on lightwave components*, Microwave and Optical technology letters, vol 5 no. 11, pp 549-553, Oct 1992.
- [902] R.F.M. van den **Brink**,: *Improved lightwave synthetic noise generator using noise injection and triangular modulation*. IEEE Photonics technology letters, vol 6, no 4, pp579-582, April 1994.
- [108] C.M. de **Blok**, R.F.M. van den **Brink**, M.J.M. van **Vaalen**, P.J.M. **Prinz**: *Fast low-cost feed-forward burst-mode optical receiver for 622 Mb/s*: EFOC & N'93 the Eleventh Annual Conference, June 30, 1993. PTT Research, NT-PU-93-1242.
- [107] **Labrujere**, **Deventer**, **Koning**, **Bekooij**, **Tan**, **Roelofsen**, **Lange**, **Boly**, **Berendschot-Aarts**, **Spruit**, **Nielen**, **Brink**, **Blok**, **Bochove**: *COSNET - a coherent optical subscriber network*. IEEE Journal of lightwave techn, vol 11, no 5/6, pp 865-874, May/June 1993.

International publications (acoustics)

- [119] C.M.de **Blok**, R.F.M. van den **Brink**: *Full characterization of linear acoustic networks based on N-ports and s-parameters*. J. Audio Eng Soc., Vol 40 no. 6, pp 517-522, June 1992.
- [120] C.M.de **Blok**, R.F.M. van den **Brink**: *Direct-reading one-port acoustic network analyzer*. J. Audio Eng Soc., Vol 41 no. 4, pp 231-238, April 1993.
- [121] R.F.M. van den **Brink**, C.M.de **Blok**: *A one-port acoustic network analyzer with full error correction*. - to be published.
- [122] R.F.M. van den **Brink**, C.M.de **Blok**: *Robust calibration methods for one-port acoustic network analyzers with multiple standard reflectors*. - to be published.

National publications

- [109] R.F.M. van den **Brink**: *The characterization and modelling of opto-electronic devices*, Tijdschrift van het NERG, vol 57 no. 3, pp 103-108 (In Dutch), 1992, PTT-Research internal report NT-PU-92-1206.
- [110] R.F.M. van den **Brink**: *Wideband print layout design; analog aspects in designing print layouts for digital circuits*. EMC in theory and practice, Holland Electronica / FME , The Netherlands, June 1993, (in Dutch) ISBN 96-71306-60-7, PTT internal: NT-PU-93-1234
- [101] R.F.M. van den **Brink**: *Broadband analogue electronics*. PTT Research Yearbook 1989, Leidschendam the Netherlands

ESPRIT reports

- [609] P.J.M. **Prinz**, R.F.M. van den **Brink**: *Report of first prototype of 900 MHz optical receiver preamplifier, deliverable 4.23*, PTT-Research, International ESPRIT 2054: UCOL report 938/90, Oct 1990.
- [610] R.F.M. van den **Brink**: *Optical receivers with capacitive current-current feedback*, ESPRIT Conference/exhibition, PTT-Research, International ESPRIT 2054: UCOL report 1139/90, Nov 1990.
- [612] P.J.M. **Prinz**, R.F.M. van den **Brink**: *Report of prototype of receiver preamplifier without tuning, deliverable 4.24*, PTT-Research, International ESPRIT 2054: UCOL report 483/91, April 1991.
- [613] E. **Drijver**, R.F.M. van den **Brink**, H.J.T. van der **Vleut**, P.J.M. **Prinz**: *Report on Tuning, deliverable 4.39*, PTT-Research, International ESPRIT 2054: UCOL report NT-INT-91-1456, Dec 1991.

PTT Research internal reports

- [606] R.F.M. van den **Brink**: *Noise Optimum of a fibre optic receiver pre-amplifier* PTT-Research, Memorandum 1391 DNL/86, Dec 1986

- [607] M.J. van der **Pol**, R.F.M. van den **Brink**: *A wideband optical receiver using third order current-current feedback* (In Dutch), PTT-Research, Memorandum 1526 DNL/87, July 1987.
- [111] R.F.M. van den **Brink**: *Handbook SABEL/CAE*. PTT-Research internal report NT-RAP-92-28, Jan 1992 (In Dutch) This is a complete overview of the system and refers to all relevant internal documents
- [112] R.F.M. van den **Brink**: *SABEL/CAE: Microwave toolbox, users manual and theoretical backgrounds version 1.0*. PTT-Research internal report NT-ER-91-1298, Dec 1991 (In Dutch)
- [113] R.F.M. van den **Brink**: *LOCUS Mathematics on some elementary polynomial routines*. PTT-Research internal report 1538DNL/87, 1987 (In Dutch)
- [114] R. **Robbertsen**, R.F.M. van den **Brink**: *LOCUS Mathematics on iteration routines for multi-variable functions (Levenberg-Marquardt iteration)*. PTT-Research internal report 1727DNL/88. (In Dutch)
- [115] R.F.M. van den **Brink**: *LOCUS user's manual version 1.5*, PTT-Research internal report 110RNL/89. 1989 (In Dutch)
- [116] T.A. van der **Laan**, R.F.M. van den **Brink**: *LOCUS algorithm for the determination of the frequency response of electrical circuits, in closed analytical form*. PTT-Research internal report 476RNL/89. 1989, (In Dutch)
- [117] J.J. van **Wamel**, R.F.M. van den **Brink**: *LOCUS a study to automated compensation*. PTT-Research internal report 830RNL/89. 1989 (In Dutch)
- [118] A.J. 't **Jong**, R.F.M. van den **Brink**: *LOCUS Analytical circuit simulation*. PTT-Research internal report 470/91. 1991 (In Dutch)

Chapter 2

Formal description of transfer

The analysis of parasitics, as discussed in section 1.2.3, is of basic importance for wideband feedback design. Designers can analyze parasitic effects when they are adequately quantified.

Quantification of parasitics requires an unambiguous definition of transfer properties of devices. The basis of an adequate transfer description, including parasitic interactions, is a proper description of wideband *signals*. This is essential since signal flow is highly affected by the geometry of the print layout on the transport of signal to the devices.

This chapter provides these theoretical tools, and starts with a detailed description of complex signal flow (magnitude and phase) in a well-defined geometry. It summarizes the formal definitions of load impedances, sources, and transfer in arbitrary multi-port networks. All discussions are focused on *linear* descriptions.

Highlights of this chapter

In this chapter, a (micro)wave approach for describing signals and devices is followed. It relies on known concepts¹ of waves, reference planes and matrix parameters to describe devices as blackboxes. All these aspects are discussed in this chapter.

The highlights of this chapter, which have been explored in this study, are:

- A detailed overview of known definitions and relations on signal flow through reference planes. It is applicable to voltages and currents as well as to waves normalized to arbitrary (complex) reference impedances.
- A detailed overview of known matrix methods for describing one-port, two-port and multi-port networks, and for describing their properties and relations. The validity of the formulae includes signal flow with arbitrary waves, each port normalized to a different and arbitrary reference impedance.
- Identification of a new class of two-port parameters, named *virtual circuit parameters*, to specify arbitrary two-port measurements in a way that is closely related to the equivalent circuit models. This property of being closely related is of significant advantage when specifying measured two-port parameters of bipolar and field effect transistors, without any accuracy limitations of models.

¹ We emphasize that sections 2.1 and most of 2.2 are essentially known microwave concepts. Nevertheless, we observed that many microwave textbooks give an incomplete overview on this subject, and have restricted the validation of the supplied formulae to special cases without further notice. Since these principles are of basic importance for the rest of this book, we start this chapter with a detailed overview of these concepts. We had to consult many textbooks and articles to complete this detailed overview.

2.1. Signal flow observed relative to reference planes

With increasing bandwidth, the dimensions of electronic circuits as well as individual components and their constituent parts become significant with respect to wavelength. In this case, many of the assumptions characteristic of relatively low frequency design lose their validity, and a more fundamental *electromagnetic* signal description increases in relevancy.

This is obvious in the extreme for radio links, antenna design and optics, where the observed dimensions are in the same order of magnitude (or larger) than the wave length of interest. In most practical situations it is however unnecessary to use a full electromagnetic signal description, although we will demonstrate that many of the concepts introduced by microwave engineers are useful to the problem at hand.

In this section 2.1, we review the use of standard microwave methods to define signal flow, matching use of voltage, current and wave methods simultaneously.

single mode transmission

Devices on printed circuit boards (PCB) are interconnected with copper conductors. When individual components are criss-cross interconnected, adequate signal flow is restricted to relatively low frequencies. At higher frequencies, cross-talk effects can be minimized by using copper structures that concentrate the electromagnetic fields in oblong volumes [110], most usually in the form of planar transmission lines above a ground plane.

When designing and characterizing wideband circuits, it is important to restrict signal flow to a single mode². This is because circuit designers prefer simplified descriptions of signal flow using voltages, currents or traveling waves rather than electromagnetic field equations. The signal frequency and the geometry³ of the guiding structure are decisive in determining whether links are single mode or not. For instance, propagation on microstrip lines (see figure 2.1) is quasi-TEM in nature. In practice the TEM approximation is valid to several GHz for 1.5 mm epoxy PCB.

Single mode signal flow facilitates the required simplification because the field in all points of a plane perpendicular to the propagation direction is calculable⁴ when the field is measured in a single point. The consequence is that, one complex number per frequency is adequate for describing forward single-mode transmission. Multi-mode links require additional information for characterizing forward signal flow, and are therefore inconvenient⁵ for transmission purposes.

² In transmission lines, more than one field pattern may fulfill the Maxwell equations as well as the boundary condition of the structure. When no more than one field pattern *propagates* in the transmission line, this link is called a *single mode* link. When more patterns propagate, it is referred as a *multi mode* link.

³ Transmission lines whose diameter is in the same order of magnitude as the wave length of interest, are often single mode transmission lines. In a poorly designed printed circuit board (PCB) with criss-cross links between the devices modes are undefined. The result is that the signal flow may become unpredictable above several MHz. A well-designed PCB with microstrip layout may improve this limit above tens of GHz. These high values usually hold for commonly used coaxial cables (0.5 cm thickness). Rectangular wave guides become multi mode one octave above the cut-off frequency. The first generation fibers (before 1983) were multi mode fibers because the technology was initially unable to handle very small core diameters

⁴ This is restricted to links with unambiguous polarization. This does not hold for structures such as circular wave guides

⁵ A graded-index (multi-mode) fiber with a 50μm core typically allows around 500 different core modes, each of which has different signal levels and propagation- and attenuation characteristics. Beyond a certain

We emphasize that a single mode link does not exclude the existence of higher order modes locally. In single mode links, higher order modes damp out quite close to the obstacle that has excited them.

The signal flow in the structure of figure 2.2 is principally single mode at the locations demarcated with solid lines. This is because excited higher order modes do not propagate in this example. The junctions in the center of figure 2.2 may excite higher order modes, which may interact in the vicinity of the dashed line.

signal flow description parameters

A single mode line has two directions of propagation: forward and reverse. Each frequency component of the propagating signal is scaled in magnitude and phase. As a result, two complex numbers per frequency are adequate to provide a full bi-directional description of single mode electromagnetic signal flow.

Voltage and current (U,I) as well as forward and reverse waves (Ψ^+ , Ψ^-) are suitable descriptive pairs for the signal propagation. When the circuit dimension exceeds the wave length of interest, a signal description using waves is commonly preferred. When designing at smaller dimensions or lower frequencies, voltages and currents are more convenient.

Both descriptive methods are identical and are convertible in each other when U and I are well-defined⁶. Both methods are inadequate at locations where higher order modes are excited. Most textbooks focus on one of these two, ignoring or underexposing the other. One aim of this chapter is to discuss them simultaneously in a more balanced way.

length of fiber (up to many kilometers) a steady-state power distribution of fiber modes can be observed to be independent of the launching conditions. Therefore they are applicable under limited conditions.

⁶ Note that the definition of voltage and current in practical links with lossy conductors and isolators is a non-trivial problem.

2.1.1. Signal definition in paired conductor links

When the dimensions of an entire printed circuit board are small, compared to the shortest wavelength of interest, the use of reference planes is of minor importance. In a short conductor the potential is position independent, and all current densities are in phase. As a result, the chosen conductor layout and the reference planes are irrelevant. In wideband circuits however the print layout is as important as the electronic circuit itself. Good design implies the use of well-defined transmission lines.

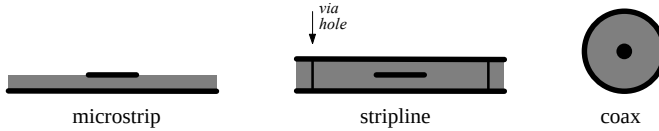


Fig 2.1 Examples of cross-sections of transmission lines. The microstrip and stripline configurations are open configurations and therefore less than ideal. Nevertheless, their fields are mainly concentrated in the dielectric medium between the two conductors which enhances their applicability. The coaxial configuration is completely closed. This prevents electromagnetic interaction with other coaxial transmission lines.

Figure 2.1 shows some examples of well-defined transmission lines, useful to interconnect components and circuits.

- The microstrip is suitable for constructing wideband circuits on printed circuit boards. The bottom side, which is metallized, acts as a ground plane while the copper traces at the top side interconnect all components.
- The stripline is suitable for wideband constructions in multi layer printed circuit board. The two ground planes are interconnected where requested using via holes. A stack of many stripline-layers forms a multi-layer print with minimal mutual coupling between the individual stripline layers.
- The coaxial structure is suitable for interconnection of different circuit blocks.

Figure 2.2 shows some examples of reference planes on a print with a microstrip layout. For a detailed discussion of designing print layout, see van den Brink [110].

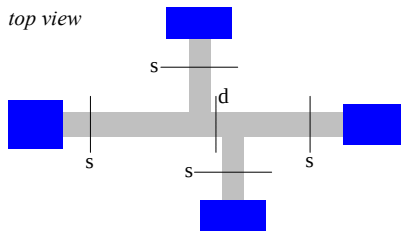


Fig 2.2 Examples of reference planes on a printed circuit board. Four components are interconnected, using a microstrip line. The solid (s) and dashed (d) lines demarcate arbitrary reference planes. At very high frequencies, e.g. above 1 GHz, the signal observed relative to the dotted reference plane may suffer from higher order modes.

Figure 2.3 shows a symbolic representation of the signal flow from a source to a load, observed relative to an arbitrary reference plane. Figure 2.3b defines the single mode waves Ψ from the voltage and the current, assuming that U and I are well-defined⁷. Figure 2.3a shows the reverse transformation.

The reference impedance Z_N is an arbitrary chosen (complex) impedance. Originally, this impedance was chosen to be the (real) characteristic impedance Z_0 of the transmission line, to study scattering effects [204,206]. In 1960, Penfield [207] introduced complex power waves for the discussion of noise performance of negative resistance amplifiers. In 1961 Youla [208] explored the same waves to simplify calculations on power flow. In 1965 Kurokawa [210,211] generalized the theory on waves associated with complex reference impedances.

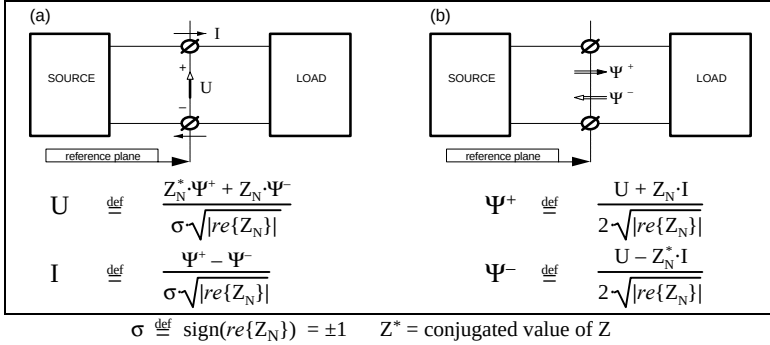


Fig 2.3 The signal flow through reference planes can be described with voltages and currents as well as with waves. The sign correction factor σ extends the validity of the definition to negative reference impedances Z_N .

The definitions, summarized in figure 2.3, are valid for any complex reference impedance, including impedances with negative real part. Many textbook discussions of generalized waves are restricted to complex impedances with positive real part. Some textbooks further restrict the discussion to real reference impedances $R_N = Z_N$. Under these restricted conditions, the signal parameters simplify to:

$$\begin{aligned} U &\rightarrow (\Psi^+ + \Psi^-) \sqrt{R_N} & \Psi^+ &\rightarrow (U + R_N \cdot I) / (2 \sqrt{R_N}) \\ I &\rightarrow (\Psi^+ - \Psi^-) / \sqrt{R_N} & \Psi^- &\rightarrow (U - R_N \cdot I) / (2 \sqrt{R_N}) \end{aligned}$$

The average power that flows through the reference plane equals $\text{re}\{U \cdot I^*\}$. This flow is sometimes referred as transported power, or as transferred power. Using the definitions of figure 2.3 for relating this power flow to waves, it can be easily demonstrated that average power flow is equal to:

$$\langle P \rangle \stackrel{\text{def}}{=} \text{re}\{U \cdot I^*\} = \sigma \cdot |\Psi^+|^2 - \sigma \cdot |\Psi^-|^2 = \text{average power flow}$$

⁷ This condition excludes waveguides and fiber-optic links. In these links, the wave concept is more appropriate.

2.1.2. Equivalent load models

For signal flow to a passive linear load, the ratio between the two signal parameters is independent of the signal source. Commonly used ratios are referred to as impedance, admittance or reflection coefficient. Their definition is shown in figure 2.4, when the observation orientation relative to the reference plane is chosen in the direction of the load. The arrows in these equations indicate the simplified definitions when the reference impedance is real $Z_N = R_N$.

load impedance	$Z \stackrel{\text{def}}{=} \frac{U}{I} = \frac{Z_N^* + Z_N \Gamma}{1 - \Gamma}$	$\rightarrow \frac{1 + \Gamma}{1 - \Gamma} \cdot R_N$
load admittance	$Y \stackrel{\text{def}}{=} \frac{I}{U} = \frac{1 - \Gamma}{Z_N^* + Z_N \Gamma}$	$\rightarrow \frac{1 - \Gamma}{1 + \Gamma} / R_N$
load reflection coefficient	$\Gamma \stackrel{\text{def}}{=} \frac{\Psi^-}{\Psi^+} = \frac{Z - Z_N^*}{Z + Z_N}$	$\rightarrow \frac{Z - R_N}{Z + R_N}$

Fig 2.4 Definitions of the load model parameters, and how they simplify for real reference impedance $Z_N = R_N$.

The reciprocal value of impedance is admittance. Similarly, a reciprocal definition for reflection coefficient is feasible, although not commonly used. Using the above definitions, the signal parameters are interrelated as follows:

$$\begin{aligned}
 U &= I / Y = \frac{Z_N^* + Z_N \Gamma}{\sigma \sqrt{\text{re}\{Z_N\}}} \cdot \Psi^+ = \frac{Z_N^* / \Gamma + Z_N}{\sigma \sqrt{\text{re}\{Z_N\}}} \cdot \Psi^- && \rightarrow (1 + \Gamma) \cdot \sqrt{R_N} \cdot \Psi^+ \\
 I &= U / Z = \frac{1 - \Gamma}{\sigma \sqrt{\text{re}\{Z_N\}}} \cdot \Psi^+ = \frac{1 / \Gamma - 1}{\sigma \sqrt{\text{re}\{Z_N\}}} \cdot \Psi^- && \rightarrow (1 - \Gamma) / \sqrt{R_N} \cdot \Psi^+ \\
 \Psi^+ &= \Psi^- / \Gamma = \frac{1 + Z_N / Z}{2 \sqrt{\text{re}\{Z_N\}}} \cdot U = \frac{Z + Z_N}{2 \sqrt{\text{re}\{Z_N\}}} \cdot I \\
 \Psi^- &= \Psi^+ \cdot \Gamma = \frac{1 - Z_N^* / Z}{2 \sqrt{\text{re}\{Z_N\}}} \cdot U = \frac{Z - Z_N^*}{2 \sqrt{\text{re}\{Z_N\}}} \cdot I
 \end{aligned}$$

The average power that is consumed (absorbed) in the load equals the average power flow through the reference plane. Its value is for arbitrary values Z_N :

$$\boxed{\langle P_L \rangle \stackrel{\text{def}}{=} |U|^2 \cdot \text{re}\{1/Z\} = |I|^2 \cdot \text{re}\{1/Y\} = \sigma \cdot |\Psi^+|^2 \cdot (1 - |\Gamma|^2)} = \text{consumed power}$$

2.1.3. Equivalent source models

For signal flow from an (active) linear source, the source can be represented by a simple equivalent model. Commonly used equivalent models are Thévenin, Norton and wave models. They are defined in figure 2.5, when the observation orientation relative to the reference plane is chosen in the direction of the load. The associated source parameters are directly extracted from two signal measurements with different (unknown) loads.

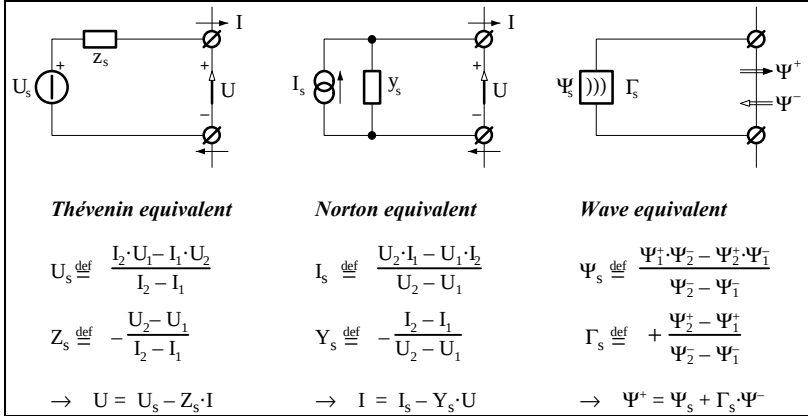


Fig 2.5 Definition of three equivalent models for linear sources, using voltage, currents and waves. The indices "1" and "2" refer to two measurements using different loads.

Many design applications give strong preference to one specific representation of a source. From a *mathematical* point of view, the equivalent models in figure 2.5 are fully transformable to each other. Figure 2.6 and 2.7 show how they are interrelated. Note that the relations between source impedance and reflection coefficient are slightly different from the relations between load impedance and reflection coefficient, shown in figure 2.4. The reference impedances are conjugated because the observation orientation relative to the reference plane is directed towards the load.

Thévenin equivalent	Norton equivalent	Wave equivalent
$U_s = U_s$	$I_s = U_s/Z_s$	$\Psi_s = \frac{\sigma \cdot \sqrt{\text{re}\{Z_N\}}}{Z_s + Z_N^*} \cdot U_s$
$U_s = I_s/Y_s$	$I_s = I_s$	$\Psi_s = \frac{\sigma \cdot \sqrt{\text{re}\{Z_N\}}}{1 + Y_s \cdot Z_N^*} \cdot I_s$
$U_s = \frac{2 \cdot \sqrt{\text{re}\{Z_N\}}}{1 - \Gamma_s} \cdot \Psi_s$	$I_s = \frac{2 \cdot \sqrt{\text{re}\{Z_N\}}}{Z_N + Z_N^* \cdot \Gamma_s} \cdot \Psi_s$	$\Psi_s = \Psi_s$
$Z_s = \frac{Z_N + Z_N^* \cdot \Gamma_s}{1 - \Gamma_s}$	$Y_s = \frac{1 - \Gamma_s}{Z_N + Z_N^* \cdot \Gamma_s}$	$\Gamma_s = \frac{Z_s - Z_N}{Z_s + Z_N^*}$

Fig 2.6 Relations between the source model parameters. The sign correction factor $\sigma \stackrel{\text{def}}{=} \text{sign}(\text{re}\{Z_N\})$ expands the validity of these models for negative reference impedances.

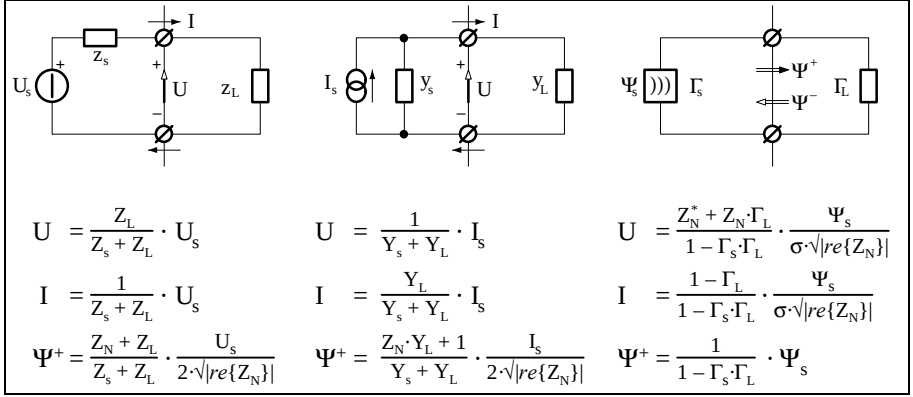


Fig 2.7 Calculation of the signal parameters U , I and Y using the equivalent models for source and load.

The exchangeable source power P_s is the stationary value of the power that will exchange between source and load by arbitrary variation of the load impedance [715,716]. P_s is:

$$\langle P_s \rangle = \frac{|U_s|^2}{4 \cdot re\{Z_s\}} = \frac{|I_s|^2}{4 \cdot re\{Y_s\}} = \sigma \cdot (1 - |\Gamma_s|^2) \cdot |\Psi_s|^2 = \text{exchangeable source power}$$

The optimum load conditions for this purpose is when $Z_L = Z_s^*$ or $\Gamma_L = \Gamma_s^*$. When $re\{Z_s\} > 0$ the maximum available power equals the exchangeable power. When $re\{Z_s\} < 0$ the maximum available power would always be infinite. Note the very subtle difference in the real parts of Z , Y and Γ between the equations for exchangeable source power and consumed power in the load, since: $re\{1/Z\} \neq 1/re\{Z\}$. This is illustrative of the need for care in extending real formulae to complex impedances.

2.1.4. Interconnection transformation of waves

Observation relative to a reference plane has an orientation, which defines the direction of forward (incident) waves (Ψ^+). When the reference impedance is real, then reversal of observation orientation is simply the reversal of forward and reverse waves. This does not hold when the reference impedances are complex.

When observing sources and loads independently, then the chosen observations directions are usually oriented *into* the networks. This makes that both observations will not match when signal flow is analyzed between interconnected networks. It requires an interconnection transformation of observed waves to undo differences in orientation, and if necessary to undo differences in reference impedances.

Figure 2.8 shows two different observations of waves, $\{\Psi_a^+, \Psi_a^-\}$ and $\{\Psi_b^+, \Psi_b^-\}$, of the same signal flow. Reordering of the wave definitions in figure 2.3 results in the following matrix relations:

$$\begin{bmatrix} \Psi_a^+ \\ \Psi_a^- \end{bmatrix} = \frac{1}{2\sqrt{re\{z_a\}}} \begin{bmatrix} 1 & +Z_a \\ 1 & -Z_a^* \end{bmatrix} \cdot \begin{bmatrix} U \\ I \end{bmatrix} \quad \begin{bmatrix} \Psi_b^+ \\ \Psi_b^- \end{bmatrix} = \frac{1}{2\sqrt{re\{z_b\}}} \begin{bmatrix} 1 & +Z_b \\ 1 & -Z_b^* \end{bmatrix} \cdot \begin{bmatrix} U \\ -I \end{bmatrix}$$

$$\begin{bmatrix} U \\ I \end{bmatrix} = \frac{1}{\sigma_a \sqrt{\operatorname{re}\{Z_a\}}} \begin{bmatrix} Z_a^* & Z_a \\ 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} \Psi_a^+ \\ \Psi_a^- \end{bmatrix} \quad \begin{bmatrix} U \\ -I \end{bmatrix} = \frac{1}{\sigma_b \sqrt{\operatorname{re}\{Z_b\}}} \begin{bmatrix} Z_b^* & Z_b \\ 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} \Psi_b^+ \\ \Psi_b^- \end{bmatrix}$$

The transformation of the waves $\{\Psi_a^+, \Psi_a^-\}$ into $\{\Psi_b^+, \Psi_b^-\}$ is simply a matrix product, extracted from the above mentioned relations. The associated *interconnection matrix* Λ is defined in figure 2.8.

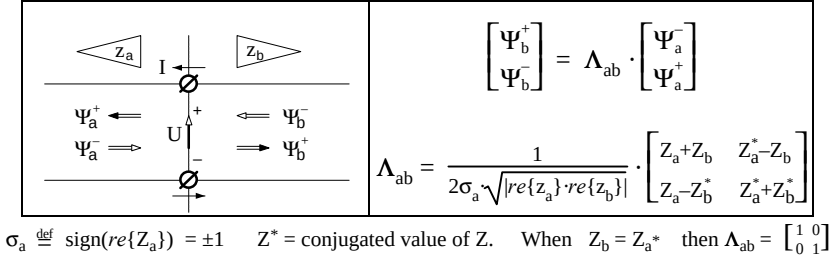


Fig 2.8 Interconnection transformation of waves observed from opposite orientation, each pair normalized to different reference impedances.

When all reference impedances are real and equal then the interconnection transformation is very simple. The interconnection matrix then equals the identity matrix. Transformation is then simplified to exchanging reflected waves with incident waves.

Occasionally, it is convenient to use complex reference impedances [210], for instance to simplify power calculations on waves generated by sources with complex output impedances. Then it is preferred to choose Z_b as the complex conjugate of the reference impedance Z_a . In all other situations the above transformation is required.

2.1.5. Conclusions

Various definitions on signal magnitude, source models and load models are summarized. Our basic philosophy of carrying out observations (measurement) relative to well-defined reference planes, while simultaneously discussing of U , I and Ψ , was introduced. As a result, no specific prior knowledge was required as to sources and loads.

Although no new definitions were introduced, many textbooks and articles have contributed to this overview, since most texts ignore waves and focus on voltages and currents, or are restricted to real impedances, or to complex impedances with positive real part.

The subtle differences between the equations for consumed power and available power demonstrate how risky it is to extend formulae for real impedances to the complex case. The same applies for the subtle differences in the relations between impedance and reflection coefficient in source and load. As will be seen, this overview lays the foundation for the rest of this work.

2.2. Blackbox representation of multi-port networks

Modeling linear⁸ devices as ideal *lumped*⁹ elements is often adequate when designing at relatively low frequencies¹⁰. In these cases, the lumped element values are successfully estimated from low frequency or dc-measurements.

Parasitic capacitive and inductive effects may degrade the validity of low-frequency models. These parasitic effects are considered as *distributed* when the device dimensions cannot be ignored compared to the shortest wavelength (highest frequency) of interest. An infinite number of lumped elements are then required to model these effects, which is considered to be impractical. A well-known example of a distributed effect is delay in transmission lines or transistors.

Wideband circuit analysis requires a mixed analysis of lumped as well as distributed elements. Lumped elements are to be considered as special distributed elements, and each distributed element can be considered as blackbox. These blackboxes are interconnected via ports. When the signal flow through each port is observed relative to well-defined reference planes, the external behavior of each blackbox can be measured at each frequency of interest. These measurements result in blackbox descriptions in tabular form that are directly applicable in microwave circuit simulators.

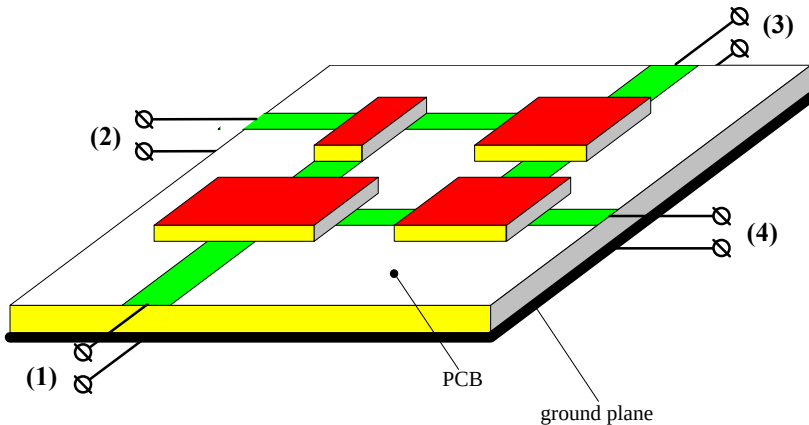


Fig 2.9 An example of a four-port blackbox, which is a sub-circuit that has been cut from a larger PCB (printed circuit board). The component leads are interconnected at the top side of the PCB or grounded using metalized via holes to a uniform copper layer at the bottom side. The edges of this PCB with microstrip layout define the reference planes.

A *linear* blackbox approach, using matrix parameters as description, is applicable over very wide frequency intervals. It remains applicable as long as the signal flow through

⁸ In this text, all *linear* networks are considered as *time invariant* linear networks.

⁹ Lumped elements are ideal elements with zero dimensions and zero delay. Examples are ideal resistors, capacitors, inductors, transformers, and voltage controlled current sources.

¹⁰ The concept 'low frequencies' is highly dependent on the circuit construction. It may be as high as 100 MHz (or higher) when using SMD (surface mounted devices) components on a microstrip print-layout, or as low as 1 MHz (or lower) when all components are at randomly connected to a print with ineffective ground planes.

the reference planes is single mode in nature. This approach in fact restricts the linear analysis to the description of external circuit behavior.

Figure 2.9 shows an example of a four-port blackbox with four well-defined reference planes. It illustrates a piece of a PCB (printed circuit board) that is part of a larger circuit, for instance an amplifier stage of a feedback amplifier with many amplifier stages. The reference planes equal the edges of this piece of PCB.

The thin microstrip construction facilitates that most of the electromagnetic activity is concentrated *in* the (glass-epoxy) dielectric of the PCB. As a result, it is assumed that the radiation is negligible and that the waves pass the reference planes as single mode waves (plane waves).

In section 2.2 we review in short known matrix methods (z-,y-,s-parameters) for the description of multi ports (blackboxes) In addition, we propose virtual circuit methods for describing the transfer of two-port, such as transistors.

2.2.1. Multi-port matrix parameters

Matrix methods are well-known for the use in describing physical multi-port networks. Voltage and current methods (z-, y-parameters) on the one hand and wave methods (s-parameters) on the other have evolved over the last sixty years.

- As early as 1929 [201] Strecker and Feldtkeller proposed these methods, and before 1940 many refinements were added to the theory, for instance [202]. As early as 1953 [205] the application of Z and Y parameters became common place in standard text books on transistor circuit design. Many textbooks have followed since, using matrix methods based on voltages and currents. Many of these texts neglect to mention the existence of S-parameters, associated with a wave description, as an alternative method.
- On the other hand, many classic texts describe wave propagation on transmission lines, using reflection and transmission coefficients. Physicists familiar with wave particle scatter problems, were active at the Radiation Laboratory of M.I.T. in the early 1940's. They applied scattering concepts to the study of wave guide junctions. A summary¹¹ appeared in 1948 by Montgomery [204], and in 1956 Carlin [206] published a detailed analysis on s-parameters¹² with many relevant references. Initially, real reference impedances were used. In 1960, Penfield [207,209] and later Youla [208] introduced s-parameters with complex reference impedances. In 1965, Kurokawa [210,211] generalized the theory, using arbitrary complex reference impedances which are port dependent.

Initially the measurement of multi-port parameters, observed relative to well-defined reference planes, was very laborious. Deschamps [212] proposed in 1953 a graphical method, using slotted lines and movable shorts, to perform full two-port parameter measurements. The introduction in 1967 of automated microwave network analyzers [213,214] simplified these measurements significantly. In the 1970's the popularity of s-parameters increased among engineers due to the growing availability of automated

¹¹ In the years 1942 to 1946 a rather intensive and systematic exploitation of microwave problems was carried out at the Radiation Laboratory of MIT. In the years 1946 to 1952 a group of MIT workers wrote 28 volumes dealing with microwave and related techniques, developed during World-War II.

¹² It was a special issue on scattering parameters, with several other papers on the same subject.

network analyzers. In recent years a growing number of manufacturers have taken to the specification of transistors in terms of s-parameters. Recently, the specification of the s-parameters of passive components (capacitors, inductors) has become more common place.

As a result, methods using z-, y- and s-parameters are now well-known. Nonetheless, no single text was found which reviewed all aspects of the relationship of z-, y-, and s-parameter description in a rigorous way.

<i>definition of symbols</i>	
$\mathbf{Z}_N$	is diagonal matrix, with arbitrary reference impedances $\{Z_{Nk}\}$
Z_0	reference impedance when they are all equal and complex
R_0	reference impedance when they are all equal and real
$\mathbf{r}_N$	is $\sqrt{\text{re}\{\mathbf{Z}_N\}}$ is a diagonal matrix with real elements
σ	is $\text{re}\{\mathbf{Z}_N\}/ \text{re}\{\mathbf{Z}_N\} $ is a diagonal matrix with ± 1 elements = $\text{inv}(\sigma)$
$\mathbf{1}$	is identity (diagonal) matrix
$\mathbf{Z}^*$	is the conjugate of matrix $\mathbf{Z}$
$\mathbf{Z}^T$	is the normal transpose of matrix $\mathbf{Z}$
$\mathbf{Z}'$	is the conjugated transpose of matrix $\mathbf{Z}$ (=adjoint matrix): $\mathbf{Z}' \equiv \mathbf{Z}^{*T}$

Fig 2.10 Definition of symbols that are used in this text. We often prefer Z' as notation for complex conjugated scalars since in that case $Z' \equiv Z^*$.

$\begin{bmatrix} U_1 \\ U_2 \\ \dots \\ U_n \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1n} \\ z_{21} & z_{22} & \dots & z_{2n} \\ \dots & \dots & \dots & \dots \\ z_{n1} & z_{n2} & \dots & z_{nn} \end{bmatrix} \cdot \begin{bmatrix} I_1 \\ I_2 \\ \dots \\ I_n \end{bmatrix}$	impedance matrix: $\mathbf{Z}$ $\langle P \rangle = \sum_k \text{re}\{U_k^* \cdot I_k\}$ $\langle P \rangle = \text{re}\{\mathbf{U}' \cdot \mathbf{I}\} = \text{re}\{\mathbf{U}' \cdot \mathbf{Z} \cdot \mathbf{U}\}$
$\begin{bmatrix} I_1 \\ I_2 \\ \dots \\ I_n \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \dots & \dots & \dots & \dots \\ y_{n1} & y_{n2} & \dots & y_{nn} \end{bmatrix} \cdot \begin{bmatrix} U_1 \\ U_2 \\ \dots \\ U_n \end{bmatrix}$	admittance matrix: $\mathbf{Y}$ $\langle P \rangle = \sum_k \text{re}\{I_k^* \cdot U_k\}$ $\langle P \rangle = \text{re}\{\mathbf{I}' \cdot \mathbf{U}\} = \text{re}\{\mathbf{I}' \cdot \mathbf{Y} \cdot \mathbf{U}\}$
$\begin{bmatrix} \Psi_{b1} \\ \Psi_{b2} \\ \dots \\ \Psi_{bn} \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{2n} \\ \dots & \dots & \dots & \dots \\ s_{n1} & s_{n2} & \dots & s_{nn} \end{bmatrix} \cdot \begin{bmatrix} \Psi_{a1} \\ \Psi_{a2} \\ \dots \\ \Psi_{an} \end{bmatrix}$	scattering matrix: $\mathbf{S}$ $\langle P \rangle = \sum_k \sigma_k \cdot \{ \Psi_{ak} ^2 - \Psi_{bk} ^2\}$ $\langle P \rangle = \Psi_a' (\sigma \mathbf{S}' \sigma \mathbf{S}) \Psi_a$
$\Psi_a = (2 \cdot \mathbf{r}_N) \cdot (\mathbf{U} + \mathbf{Z}_N \cdot \mathbf{I})$ $\Psi_b = (2 \cdot \mathbf{r}_N) \cdot (\mathbf{U} - \mathbf{Z}_N' \cdot \mathbf{I})$	$\mathbf{U} = (\sigma \cdot \mathbf{r}_N) \cdot (\mathbf{Z}_N' \cdot \Psi_a + \mathbf{Z}_N \cdot \Psi_b)$ $\mathbf{I} = (\sigma \cdot \mathbf{r}_N) \cdot (\Psi_a - \Psi_b)$

Fig 2.11 Definition of the elementary matrix parameters, and their associated power consumption relations. The symbols are defined in figure 2.10. The orientation of all observations relative to reference planes is directed into the multi-port. It means that the current flow polarity I and the 'approaching' waves Ψ_a are directed into the multi-port. All reference planes may have different reference impedances Z_N to define the waves from the voltages and currents.

In figure 2.10 and 2.11, elementary matrix parameters are defined. They are convenient in combination with modern matrix tools (e.g. Matlab[®] [123]) or modern circuit simulators for handling measured data on multi-port networks.

Z-parameters are convenient in combination with serial impedances and voltage sources). Y-parameters are convenient in combination with shunt admittances and current sources. S-parameters are convenient in combination with waves and scattering problems.

Figure 2.12 shows how these matrix descriptions are related to each other. If desired, the reference impedances Z_N on each port may be chosen differently. This is convenient, for instance, when using transmission lines with 50Ω characteristic impedance and with 75Ω . It means that $\mathbf{Z}_N$ is a diagonal matrix with all (different) reference impedances on its main diagonal.

originating from $\mathbf{Z}$	originating from $\mathbf{Y}$	originating from $\mathbf{S}$
$\mathbf{Z} = \mathbf{Z}$ $\mathbf{Y} = \text{inv}(\mathbf{Z})$ $\mathbf{S} = \mathbf{r}_N \backslash ((\mathbf{Z} - \mathbf{Z}_N') / (\mathbf{Z} + \mathbf{Z}_N)) \cdot \mathbf{r}_N$	$\mathbf{Z} = \text{inv}(\mathbf{Y})$ $\mathbf{Y} = \mathbf{Y}$ $\mathbf{S} = \mathbf{r}_N \backslash ((1 - \mathbf{Z}_N' \cdot \mathbf{Y}) / (1 + \mathbf{Z}_N \cdot \mathbf{Y})) \cdot \mathbf{r}_N$	$\mathbf{Z} = \mathbf{r}_N \cdot ((1 - \mathbf{S}) \backslash (\mathbf{S} \cdot \mathbf{Z}_N + \mathbf{Z}_N')) / \mathbf{r}_N$ $\mathbf{Y} = \mathbf{r}_N \cdot ((\mathbf{S} \cdot \mathbf{Z}_N + \mathbf{Z}_N') \backslash (1 - \mathbf{S})) / \mathbf{r}_N$ $\mathbf{S} = \mathbf{S}$
<i>all different real ref. imp.</i> $\mathbf{Z} = \mathbf{r}_N \cdot ((1 - \mathbf{S}) \backslash (\mathbf{S} + 1)) \cdot \mathbf{r}_N$ $\mathbf{Y} = \mathbf{r}_N \backslash ((\mathbf{S} + 1) \backslash (1 - \mathbf{S})) \cdot \mathbf{r}_N$ $\mathbf{S} = (\mathbf{r}_N \backslash \mathbf{Z} - \mathbf{r}_N) / (\mathbf{r}_N \backslash \mathbf{Z} + \mathbf{r}_N)$ $\mathbf{S} = (1 - \mathbf{r}_N \cdot \mathbf{Y} \cdot \mathbf{r}_N) / (1 + \mathbf{r}_N \cdot \mathbf{Y} \cdot \mathbf{r}_N)$	<i>equal complex ref. imp</i> $\mathbf{Z} = (1 - \mathbf{S}) \backslash (\mathbf{S} \cdot \mathbf{Z}_0 + \mathbf{Z}_0^* \cdot 1)$ $\mathbf{Y} = (\mathbf{S} \cdot \mathbf{Z}_0 + \mathbf{Z}_0^* \cdot 1) \backslash (1 - \mathbf{S})$ $\mathbf{S} = (\mathbf{Z} - \mathbf{Z}_0^* \cdot 1) / (\mathbf{Z} + \mathbf{Z}_0 \cdot 1)$ $\mathbf{S} = (1 - \mathbf{Z}_0^* \cdot \mathbf{Y}) / (1 + \mathbf{Z}_0 \cdot \mathbf{Y})$	<i>equal real ref. imp.</i> $\mathbf{Z} = ((1 - \mathbf{S}) \backslash (\mathbf{S} + 1)) \cdot \mathbf{R}_0$ $\mathbf{Y} = ((\mathbf{S} + 1) \backslash (1 - \mathbf{S})) / \mathbf{R}_0$ $\mathbf{S} = (\mathbf{Z} - \mathbf{R}_0 \cdot 1) / (\mathbf{Z} + \mathbf{R}_0 \cdot 1)$ $\mathbf{S} = (1 - \mathbf{R}_0 \cdot \mathbf{Y}) / (1 + \mathbf{R}_0 \cdot \mathbf{Y})$

Fig 2.12 Transformation rules between the matrix parameters of an arbitrary multiport. The validity of these equations includes reference impedances with negative real parts. These rules simplify when all ports use equal reference impedances or when they are real. See figure 2.10 for the definition of the symbols. The matrix division symbols are shortcuts, according to Matlab^a [123] conventions, for:

$$\mathbf{A} \backslash \mathbf{B} \equiv \text{inv}\{\mathbf{A}\} \cdot \mathbf{B} \quad \mathbf{A} / \mathbf{B} \equiv \mathbf{A} \cdot \text{inv}\{\mathbf{B}\}$$

2.2.2. Multi-port matrix properties

An equivalent matrix description facilitates a full description for arbitrary linear network configurations. This includes active networks with controlled sources, such as voltage controlled current sources. As a result, an arbitrary N-port network requires N^2 parameters for a full equivalent description.

For a wide range of passive networks, these parameters are interrelated. Well-known physical properties, such as reciprocal and loss-free, result in properties that simplify the matrix. This subsection summarizes these matrix conditions, since they are useful for multi-port analysis.

Reciprocal conditions

Consider an arbitrary network that is completely constructed from reciprocal materials. For a definition of reciprocal media, see [216]. Media for which the conductivity (σ),

permittivity (ϵ) and permeability (μ) are *scalar* quantities at all locations *inside* the network are reciprocal. This does not exclude position-dependency, however, these scalars must be time-invariant and independent of the amplitude and direction of the electric and magnetic field vectors $\underline{\mathbf{E}}$ and $\underline{\mathbf{H}}$.

Let us perform two arbitrary electromagnetic experiments, one that generates the fields $\underline{\mathbf{E}}^{(1)}$ and $\underline{\mathbf{H}}^{(1)}$, and another that generates the field $\underline{\mathbf{E}}^{(2)}$ and $\underline{\mathbf{H}}^{(2)}$. Using the harmonic Maxwell equations¹³ and the Gauss theorem to convert a volume integral to a closed surface integral, it can be demonstrated [211] that the following condition holds:

$$\oint\oint (\underline{\mathbf{E}}^{(1)} \times \underline{\mathbf{H}}^{(2)} - \underline{\mathbf{E}}^{(2)} \times \underline{\mathbf{H}}^{(1)}) \cdot d\Omega \equiv 0$$

When the closed surface of this integral encloses the volume of the multi-port network, this identity provides us a comparative condition for the voltages and currents at all ports. Let's restrict the experiments to a single current source and many voltage probes, one experiment that injects the current in port k and another in port l . Then it can be demonstrated [211] that the reciprocal condition is equivalent with $\{U_k^{(1)} \cdot I_k^{(2)} \equiv U_l^{(2)} \cdot I_l^{(1)}\}$ for any port number k and l . As a result, it can be demonstrated [210,211] that this condition yields the multi-port matrix properties that are summarized in figure 2.13a.

Lossless conditions

In a lossless network the total dissipated power is zero. It means that all power that enters into one port will leave the network via all ports. Integration of the average power flow $re\{\underline{\mathbf{E}} \times \underline{\mathbf{H}}^*\}$ over the closed surface of the network yields for the average power consumption:

$$\oint\oint re\{\underline{\mathbf{E}} \times \underline{\mathbf{H}}^*\} \cdot d\Omega \equiv 0 \quad \text{average power consumption}$$

It can be demonstrated [211] that this condition is equivalent with $re\{\sum_k U_k \cdot I_k^*\} \equiv 0$. This condition yields the multi-port matrix properties that are summarized in figure 2.13b.

(a) <i>reciprocal condition</i> <i>normal transpose</i>	(b) <i>loss-free condition</i> <i>conjugated transpose</i>	(c) <i>reciprocal & loss-free</i> <i>conjugated elements</i>
$\mathbf{Z} \equiv \mathbf{Z}^T \rightarrow z_{kl} \equiv z_{lk}$	$\mathbf{Z} \equiv -\mathbf{Z}'$	$\mathbf{Z} \equiv -\mathbf{Z}^* \equiv j \cdot im\{\mathbf{Z}\}$
$\mathbf{Y} \equiv \mathbf{Y}^T \rightarrow y_{kl} \equiv y_{lk}$	$\mathbf{Y} \equiv \mathbf{Y}'$	$\mathbf{Y} \equiv -\mathbf{Y}^* \equiv j \cdot im\{\mathbf{Y}\}$
$\mathbf{S} \equiv \sigma \cdot \mathbf{S}^T \cdot \sigma \rightarrow s_{kl} \equiv \pm s_{lk}$	$\mathbf{S} \equiv inv(\sigma \cdot \mathbf{S}' \cdot \sigma)$	$\mathbf{S} \equiv inv(\mathbf{S}^*)$

Fig 2.13 Matrix properties for reciprocal, for loss-free and for combined reciprocal loss-free networks. The diagonal matrix $\mathbf{S}$ represents the sign (± 1) of the chosen reference impedance Z_{nk} at each port. (See figure 2.10 for the definition).

Lossless reciprocal conditions

Networks constructed from (quasi) perfect conductors and insulators are (quasi) lossless as well as reciprocal networks. This is because the field strength ($\underline{\mathbf{E}}$, $\underline{\mathbf{H}}$) is zero in perfect conductors and because the conductivity is zero in dielectric media. Similarly to the reciprocal condition it can be demonstrated that this construction is reciprocal as well as

¹³ Maxwell equations: $\nabla \times \underline{\mathbf{H}} = (\sigma + j\omega\epsilon) \cdot \underline{\mathbf{E}}$ and $\nabla \times \underline{\mathbf{E}} = (-j\omega\epsilon) \cdot \underline{\mathbf{H}}$, Gauss: $\iiint (\nabla \cdot \underline{\mathbf{F}}) \cdot dV = \iiint \underline{\mathbf{F}} \cdot d\Omega$ and vector identities: $\nabla \cdot (\underline{\mathbf{A}} \times \underline{\mathbf{B}}) = \underline{\mathbf{B}} \cdot (\nabla \times \underline{\mathbf{A}}) - \underline{\mathbf{A}} \cdot (\nabla \times \underline{\mathbf{B}})$

loss-free. As a result, it can be demonstrated [211] that the combination of the matrix identities in figure 2.13a and 2.13b yields the additional relations of figure 2.13c. Other properties, such as symmetry or asymmetry between two ports will result in additional matrix identities, which are beyond the scope of this short overview.

2.2.3. Multi-port matrix reduction algorithm

Consider a circuit with N_n nodes, represented by an $N_n \times N_n$ matrix (z-, y- or s-parameters). In many circuit applications, no more than a few nodes ($N_p < N_n$) are externally available, for instance nodes p_1 and p_2 (the port nodes). This means that the current vector $\mathbf{I}$ has many zero coefficients except at row p_1 and p_2 . Furthermore, it means that most coefficients of the voltage vector $\mathbf{U}$ are not of interest, except at row p_1 and p_2 . This ascertainment provides the means to reduce the matrix dimension from $N_n \times N_n$ down to $N_p \times N_p$.

When the matrix representation is transformed into z-parameters ($\mathbf{U} = \mathbf{Z} \cdot \mathbf{I}$) most coefficients of $\mathbf{Z}$ are multiplied with the zero coefficients in the current vector $\mathbf{I}$. Furthermore, most calculated coefficients in the voltage vector $\mathbf{U}$ are not of interest. As a result, most coefficients in $\mathbf{Z}$ are unused and the reduced impedance matrix $\mathbf{z}$ results from $\mathbf{Z}$ by deleting all rows and columns except those with number p_1 and p_2 .

This procedure of row and column deletion is simply described using an auxiliary matrix $\mathbf{P}$ (port definition matrix) of the following general form:

$$\mathbf{P} \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{array}{l} \leftarrow \text{row } p1 \\ - \\ - \\ \leftarrow \text{row } p2 \\ - \\ - \\ - \\ - \end{array} \quad \Rightarrow \quad \boxed{\mathbf{z} = \mathbf{P}' \cdot \mathbf{Z} \cdot \mathbf{P}} \quad \text{and} \quad \mathbf{P}' \cdot \mathbf{P} = \mathbf{1}$$

This simple matrix identity is reserved for z-parameters only. Reduction of y- and s-parameters requires z-parameters as intermediate representation format. Use the transformation rules of figure 2.12 for this purpose, for instance: $\mathbf{y} = \text{inv}(\mathbf{P}' \cdot \mathbf{Y} \cdot \mathbf{P})$.

Reducing y- and s-parameters by deleting rows and columns, is equivalent with connecting the reduced nodes to ground via an impedance $Z=0$ or $Z=Z_N$ respectively.

This reduction algorithm is important for extracting various feedback parameters from arbitrary circuits in which the feedback network is specified as four-port and the forward amplifier as two-port (see section 4.2). The multi-port parameters of these sub-circuits are derived from arbitrary circuits using a circuit simulator, for instance the simulator described in section 2.2.4 and 7.2.4.

Moreover, this algorithm is important for deriving fundamental network properties, such as the generalized thermal noise theorem described in section 7.2.5.

2.2.4. Application of multi-port transfer matrices in circuit simulators

A very simple circuit simulator can be realized when using y-parameters because admittance matrix parameters are convenient when shunting circuits. A complex circuit is simply the shunted combination of many sub-circuits, each of them described with y-parameters. The *nodal admittance matrix method* described here can be found in various textbooks dealing with computer aided circuit analysis, for instance Dobrowolski [724].

Because our integrated approach of designing and characterizing wideband circuits relies extensively on various computer algorithms, this section discusses algorithms for automated circuit analysis. It facilitates the exploration of new feedback analysis and synthesis concepts, as introduced in chapter 4 and 5.

Consider two circuits, represented by $\mathbf{Y}_1$ and $\mathbf{Y}_2$. When the node voltages of these circuits equal the voltages $\mathbf{U}_1$ and $\mathbf{U}_2$, the associated currents equal $\mathbf{I}_1 = \mathbf{Y}_1 \cdot \mathbf{U}_1$ and $\mathbf{I}_2 = \mathbf{Y}_2 \cdot \mathbf{U}_2$. In the special case that all nodes are interconnected then $\mathbf{U}_1 = \mathbf{U}_2$. This is associated with a combined current flow of $(\mathbf{I}_1 + \mathbf{I}_2) = (\mathbf{Y}_1 + \mathbf{Y}_2) \cdot \mathbf{U}$. As a result, the admittance matrix of the combined circuit is simply the sum of the individual $\mathbf{Y}$ -matrices.

A simple circuit simulator [116,118] uses this property to construct the admittance matrix of complex circuits. All node numbers of the circuit are associated with a corresponding (unique) row and column number of this matrix. The ground node is zero, and corresponds with a (dummy) row and column number zero. This row and column, will be deleted afterwards. The simulator starts from an empty circuit ($\mathbf{Y} = \mathbf{0}$), and a distinct frequency of interest. Each circuit element (impedance, VCCS, etc.) is subsequently added to $\mathbf{Y}$ by calculating its individual y-parameters. This procedure is summarized below:

- An impedance Z , connected from node n_1 to n_2 , is represented by a 2×2 admittance matrix. Its four coefficients are added to four coefficients of the main $\mathbf{Y}$ matrix, that corresponds with the specified node numbers. This addition equals with:

$$\begin{bmatrix} y_{n1,n1} & y_{n1,n2} \\ y_{n2,n1} & y_{n2,n2} \end{bmatrix} = \begin{bmatrix} y_{n1,n1} & y_{n1,n2} \\ y_{n2,n1} & y_{n2,n2} \end{bmatrix} + 1/z \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad n1 \bullet \text{---} \boxed{Z} \text{---} \bullet n2$$

- A VCCS (voltage controlled current source) is essentially a four-port and must be represented by a 4×4 admittance matrix. Since it is a floating circuit, most coefficients are zero. No more than four non-zero coefficients are left, and they are added to the associated coefficients in the main $\mathbf{Y}$ matrix. This addition is:

$$\begin{bmatrix} y_{n2,n1} & y_{n2,n3} \\ y_{n4,n1} & y_{n4,n3} \end{bmatrix} = \begin{bmatrix} y_{n2,n1} & y_{n2,n3} \\ y_{n4,n1} & y_{n4,n3} \end{bmatrix} + g \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \begin{array}{c} n1 \bullet \text{---} + \\ \text{---} g \text{---} \bullet n2 \\ n3 \bullet \text{---} - \\ \text{---} \bullet n4 \end{array}$$

- A transmission line is most conveniently modeled in s-parameters. These s-parameters are converted into y-parameters, and then added to $\mathbf{Y}$.
- Most other circuit elements, such as transistor models and opamps are modeled with a combination of impedances and VCCS elements.

When all elements of the net-list are inserted, the combined admittance matrix $\mathbf{Y}$ of the circuit is complete. The inverse matrix provides the impedance matrix $\mathbf{Z}$ of the circuit. The circuit simulator must now reduce this matrix to a smaller matrix, for instance a 2×2 matrix $\mathbf{z}$ of a two-port.

As described in section 2.2.3, the reduces impedance matrix equals:

$$\boxed{\mathbf{z} = \mathbf{P}' \cdot \mathbf{Z} \cdot \mathbf{P} = \mathbf{P}' \cdot (\mathbf{Y} \backslash \mathbf{P})}$$

The circuit simulator has completed when all $\mathbf{z}$ matrices are evaluated for all frequencies of interest. When s- or y-parameters are preferred, convert all $\mathbf{z}$ matrices into the preferred format. The dimension of matrix $\mathbf{z}$ depends on the number of ports of the circuit.

We realized the above simulator in the 4GL computer language MatLab® [123], and found that less than four pages source code are adequate for implementing this simulator

2.2.5. Conclusions

Standard matrix methods for the representation of linear networks have been summarized. The overview has focused on a simultaneous discussion of z-, y- and s-parameters, without any restrictions on port reference impedance (Z_N).

Although these methods are well-known, most textbook discussions are restricted to real reference impedances, to positive complex reference impedances or to reference impedances that are equal at all ports.

Matrix expressions have been discussed for transforming N-port parameters into M-port parameters ($M < N$). The matrix notation of these expressions is convenient for proving various properties of arbitrary networks and for evaluating transfer functions in circuit simulators.

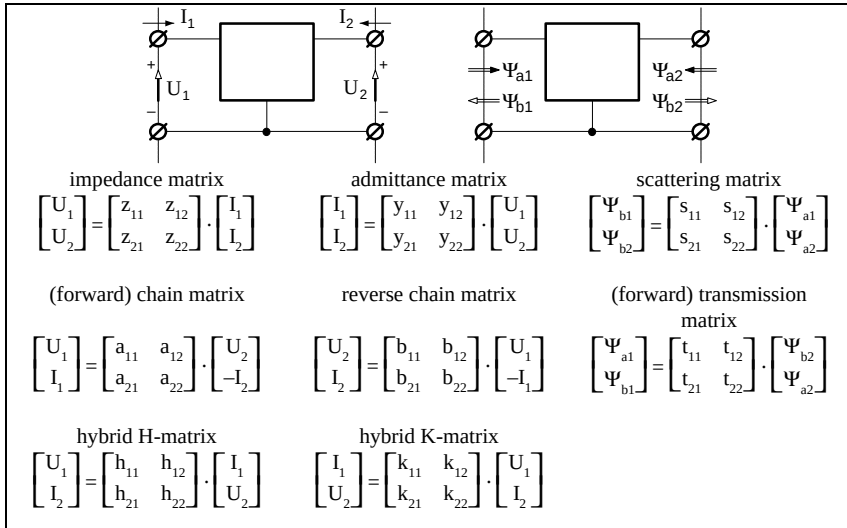
2.3. Two-port transfer parameters

An important group of multi-port parameters are two-port parameters. One of the reasons is that they are directly available from measurements with commercially available network analyzers. As a result, an increasing number of manufacturers is specifying their components (transistors, inductors, etc.) in two-port parameters, usually in s-parameter format.

This section discusses the well-known two-port matrix parameters and introduces a new class of parameters: *virtual circuit parameters*.

2.3.1. Two-port matrix parameters

In addition to previously defined z-, y- and s-parameters, many other formats are in use. This is because reorganization of the matrix equations may simplify the calculation of cascaded two-ports, or may simplify other specialized calculations. Unfortunately, different names and letters are in use for identical parameter formats. Figure 2.14 shows an overview of some preferred¹⁴ matrix parameter formats.



$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1/\mu & 1/\gamma \\ 1/\zeta & 1/\alpha \end{bmatrix} = \begin{bmatrix} 1/T_{uu} & 1/T_{iu} \\ 1/T_{ui} & 1/T_{ii} \end{bmatrix}$$

$$\begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$$

Fig 2.14 Matrix definitions for two-port networks. The definitions inside the borderline are recommended by IEC standards [215] or commonly used in microwave literature. The other matrices are equivalent definitions that are also often used.

¹⁴ The matrix symbols Z, Y, A, B, H and K are recommended by IEC standards [215] (no 3-506 to 3-511). The s- and t-parameters are not mentioned in this standard, nor the commonly used [A,B,C,D] parameters.

Analyzing a cascade of two-ports is simple when chain parameters (**A**) or transmission parameters (**T**) are used. The two-port parameters of the cascade then result from a simple matrix product of the individual two-ports matrices. These products are:

$$\begin{aligned} \text{voltage and currents:} \quad \mathbf{A} &= \mathbf{A}_1 \cdot \mathbf{A}_2 \cdot \mathbf{A}_3 \cdot \dots \mathbf{A}_n \\ \text{waves:} \quad \mathbf{T} &= \mathbf{T}_1 \cdot \mathbf{\Lambda}_{12} \cdot \mathbf{T}_2 \cdot \mathbf{\Lambda}_{23} \cdot \mathbf{T}_3 \cdot \mathbf{\Lambda}_{34} \cdot \dots \mathbf{T}_n \end{aligned}$$

The matrices **Λ** represent the interconnection transformation matrix of waves, when reference plane orientation and reference impedance do not match. This is discussed in detail in section 2.1.4. When they do match, for instance because all reference impedances are real and equal, then an interconnection transformation is superfluous because $\mathbf{\Lambda} \equiv \mathbf{I}$.

With the exclusion of a very limited number of exceptional cases, two-port matrix parameter descriptions are mutually transformable. When the reference impedances are positive real, then these relations simplify significantly. Figure 2.15 summarizes some of these relations in matrix format. Appendix A summarizes similar relations between the elements of the matrices.

$\mathbf{Z} = (\mathbf{Y}\mathbf{I})$ $\mathbf{Z} = (\mathbf{A}\cdot\mathbf{I}_3 - \mathbf{I}_1) \setminus (\mathbf{I}_2 + \mathbf{A}\cdot\mathbf{I}_4)$ $\mathbf{Z} = (\mathbf{H}\cdot\mathbf{I}_4 - \mathbf{I}_1) \setminus (\mathbf{I}_4 - \mathbf{H}\cdot\mathbf{I}_1)$ $\mathbf{Z} = \mathbf{r}_N \cdot ((\mathbf{I} - \mathbf{S}) \setminus (\mathbf{I} + \mathbf{S})) \cdot \mathbf{r}_N$	$\mathbf{Y} = (\mathbf{Z}\mathbf{I})$ $\mathbf{Y} = (\mathbf{I}_2 + \mathbf{A}\cdot\mathbf{I}_4) \setminus (\mathbf{A}\cdot\mathbf{I}_3 - \mathbf{I}_1)$ $\mathbf{Y} = (\mathbf{I}_4 - \mathbf{H}\cdot\mathbf{I}_1) \setminus (\mathbf{H}\cdot\mathbf{I}_4 - \mathbf{I}_1)$ $\mathbf{Y} = \mathbf{r}_N \setminus ((\mathbf{I} + \mathbf{S}) \setminus (\mathbf{I} - \mathbf{S})) / \mathbf{r}_N$	$\mathbf{S} = (\mathbf{r}_N \setminus \mathbf{Z} - \mathbf{r}_N) / (\mathbf{r}_N \setminus \mathbf{Z} + \mathbf{r}_N)$ $\mathbf{S} = (\mathbf{I} - \mathbf{r}_N \cdot \mathbf{Y} \cdot \mathbf{r}_N) / (\mathbf{I} + \mathbf{r}_N \cdot \mathbf{Y} \cdot \mathbf{r}_N)$ $\mathbf{S} = (\mathbf{T}\cdot\mathbf{I}_3 - \mathbf{I}_2) \setminus (\mathbf{I}_1 - \mathbf{T}\cdot\mathbf{I}_4)$
$\mathbf{A} = (\mathbf{I}_1 \cdot \mathbf{Z} + \mathbf{I}_2) / (\mathbf{I}_3 \cdot \mathbf{Z} - \mathbf{I}_4)$ $\mathbf{A} = (\mathbf{I}_1 + \mathbf{I}_2 \cdot \mathbf{Y}) / (\mathbf{I}_3 - \mathbf{I}_4 \cdot \mathbf{Y})$ $\mathbf{A} = (\mathbf{I}_1 \cdot \mathbf{H} + \mathbf{I}_2) / (\mathbf{I}_3 - \mathbf{I}_4 \cdot \mathbf{H})$	$\mathbf{H} = (\mathbf{I}_1 \cdot \mathbf{Z} + \mathbf{I}_4) / (\mathbf{I}_1 + \mathbf{I}_4 \cdot \mathbf{Z})$ $\mathbf{H} = (\mathbf{I}_1 + \mathbf{I}_4 \cdot \mathbf{Y}) / (\mathbf{I}_1 \cdot \mathbf{Y} + \mathbf{I}_4)$ $\mathbf{H} = (\mathbf{A}\cdot\mathbf{I}_4 + \mathbf{I}_1) \setminus (\mathbf{A}\cdot\mathbf{I}_3 - \mathbf{I}_2)$	$\mathbf{T} = (\mathbf{I}_1 + \mathbf{I}_2 \cdot \mathbf{S}) / (\mathbf{I}_3 \cdot \mathbf{S} + \mathbf{I}_4)$

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{I}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \mathbf{I}_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \quad \mathbf{I}_3 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \mathbf{I}_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{r}_N = \begin{bmatrix} \sqrt{Z_{N1}} & 0 \\ 0 & \sqrt{Z_{N2}} \end{bmatrix}$$

Fig 2.15 Some transformation rules between various two-port matrices when the two (different) reference impedances are positive real.

Arbitrary multi-port networks that include one or more sources are described in section 7.2 when discussing noise. The review in this chapter is restricted to networks in which a single embedded source is connected to an output port via two-port networks. These configurations are shown in figure 2.16.

When these two-port parameters are known then the internal characteristics of this source are calculable from the external values, and visa-versa. This procedure is equivalent to repositioning a reference plane from an external accessible position to an internal position, or reversal. Figure 2.16 summarizes an overview of transformation rules for embedding and de-embedding of sources.

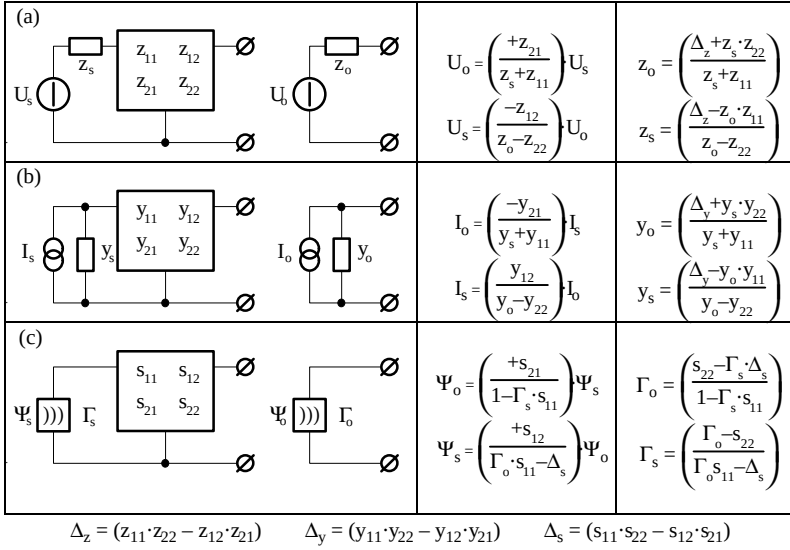


Fig 2.16 Embedding and de-embedding relations for sources.

2.3.2. Two-port virtual circuit parameters

Matrix parameters are convenient for performing two-port network calculations with computers and for representing transfer measurements. Four (measured) parameters per frequency are adequate to represent a linear two-port in tabular form. For instance, an analysis for 50 frequencies requires a table with 200 complex numbers.

Some matrix representations are more convenient for a specific job, Z-parameters are convenient when networks are connected in series, y-parameters when networks are in shunt, and s-parameters when signals are transported through transmission lines. This has caused that network analyzers usually measure s-parameters.

This is not the only matrix way for representing data from two-port measurements. With the exclusion of numerical round-off errors and a very limited number of exceptional cases, it does not really matter whether the two-port measurements are presented in s-, t-, h-, z-, a- or y-format, because all these two-port representations are mutually transformable.

Another well-known approach of describing two-ports are equivalent circuit models. These circuits are constructed with a finite number of ideal lumped elements to model the two-port. This approach facilitates an *approximate* description, and has in essence a physical meaning. The meaning of matrix parameters is in essence abstract, which facilitates an *exact* description for arbitrary linear two-ports.

We propose the identification of an additional class of two-port parameters: *virtual*¹⁵ *circuit parameters*. Similarly to other matrix parameter sets they are nonphysical in

¹⁵ We chose the name *virtual* circuits to designate it from *equivalent* circuits. Equivalent circuits usually originates on physical grounds while the meaning of virtual circuits is in essence abstract.

nature although their meaning is closely related to a simple behavioral model. For instance, amplifier blocks are often represented by an input impedance, an output impedance, a controlled source to represent gain and a controlled source to represent isolation. It means that it is characterized with a virtual circuit in stead of an equivalent circuit model. In general, virtual circuits have no physical relation to the actual amplifier. Virtual circuit parameters represent the elements of a pre-defined virtual circuit

An important aspect of virtual circuit parameters is that they are reversibly transformable from standard matrix descriptions. To enable a reverse transform two-port virtual circuits consist of exactly four independent elements. This restriction is essential for virtual circuits, while equivalent circuit models may contain any number of elements.

The advantage of two-port virtual circuit parameters is that they provide an exact description of two-ports while they have been found more appropriate for manual interpretation of two-port data than matrix parameters

Two-port virtual circuit parameters are a set of four complex numbers per frequency. Electronic circuit models, constructed with exactly four *independent* elements, are suitable for this purpose. There are many possible circuit topologies that may represent an arbitrary two-port with four independent (virtual) elements. Figure 2.17 shows some examples, using a single controlled current source and three impedances.

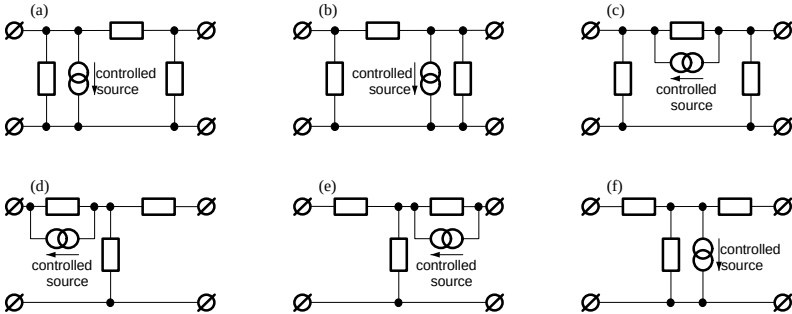


Fig 2.17 Some virtual circuit topologies that facilitate a full representation of an arbitrary linear two-port. The controlled sources are controlled by a voltage over an impedance or by a current through an impedance.

When the four-element topology is chosen, then the virtual circuit parameters can be found simply from a given matrix description. Y-parameters are often convenient for Π -topologies and z-parameters are often convenient for T-topologies. There is a reversible parameter transformation, for each frequency of interest. When the two-port is reciprocal, three virtual elements are adequate, since the matrix parameters are interrelated (see figure 2.13).

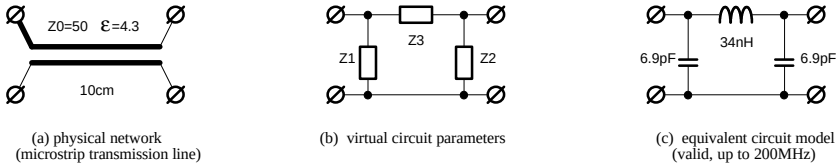


Fig 2.18 An application example of virtual circuit parameters. They simplify the validity assessment of lumped approximations of distributed networks. It can be demonstrated that the following conditions hold:
 $z_1=1/(y_{11}-y_{12})$, $z_2=1/(y_{22}+y_{12})$, $z_3=1/(-y_{12})$.

The use of virtual circuit parameters often facilitates the assessment of a lumped approximation to a distributed network. Figure 2.18a shows a well-known example of a distributed network, namely a microstrip transmission line on a standard epoxy print. The s-parameters are calculated or measured, and subsequently transformed to y-parameters to simplify the evaluation of the virtual circuit parameters.

Figure 2.18b shows the chosen virtual circuit topology. Since the two-port is reciprocal, three virtual circuit parameters are adequate. All virtual impedances Z_1 , Z_2 and Z_3 are frequency dependent.

Bode plots of Z_1 and Z_3 indicate that this frequency dependence is proportional to $(1/j\omega)$ and $(j\omega)$ respectively, over a wide frequency interval. This is confirmed by plotting $1/(j\omega \cdot Z_1)$ and $(Z_3/j\omega)$, as is shown in figure 2.19. It illustrates that Z_1 is capacitive up to 400 MHz and that Z_3 is inductive up to 200 MHz. As a result, the equivalent circuit in figure 2.18c is an adequate lumped element approximation up to 200 MHz.

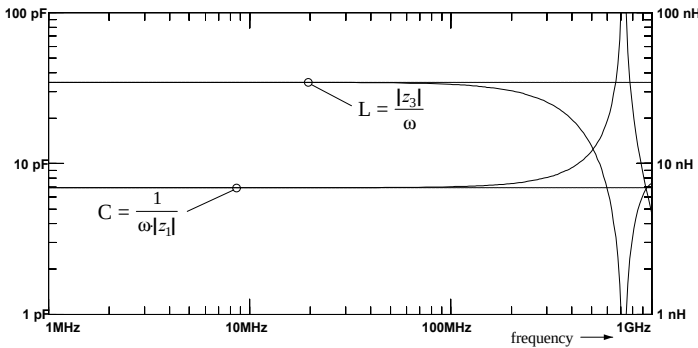


Fig 2.19 Bode plots of $1/(j\omega \cdot Z_1)$ and $(Z_3/j\omega)$, associated with the virtual circuit parameters Z_1 and Z_3 of the two-port in figure 2.18. The curves are flat over a wide frequency interval, indicating that Z_1 is capacitive up to 400 MHz and Z_3 is inductive up to 200 MHz.

Virtual circuit parameters provide an attractive graphical specification method for transistors. For discrete transistors, they can be extracted from s-parameter measurements at bias conditions of interest. For integrated transistors they must be extracted from sophisticated (non linear) device models that have various process-dependent parameters as input. More and more transistor manufacturers are also

specifying their discrete transistors in two-port parameters, usually s-parameters (tabular and graphical). These presentation formats are convenient for use with software tools and in 50Ω transmission line applications, although they are often inconvenient in other applications.

We have found plots of representative data, presented in a form that is closely related to a simple transistor model, to be preferable. This is quite different from specifying the parameters of a model, since models are approximations and virtual circuit parameters not. Finding the most convenient virtual circuit topology is a matter of trial and error in practice. A good choice is a topology in which the four virtual elements can be approximated by ideal lumped elements over the frequency interval of interest.

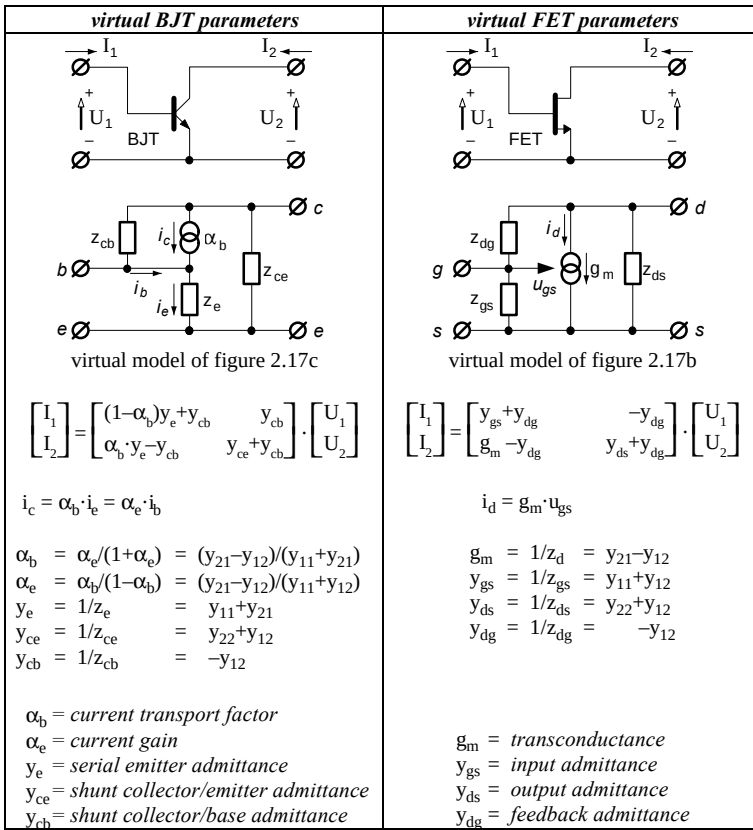


Fig 2.20 Definition of two-port virtual circuit parameters for transistors.

These parameters are $[\alpha_e, z_e, z_{ce}, z_{cb}]$ or $[\alpha_b, z_e, z_{ce}, z_{cb}]$ for a BJT, and $[g_m, z_{gs}, z_{ds}, z_{dg}]$ for an FET.

Figure 2.20 shows convenient topologies for BJT's (bipolar junction transistors) and FETs (field effect transistors). They provide the virtual parameter sets $[\alpha_e, z_e, z_{ce}, z_{cb}]$ or $[\alpha_b, z_e, z_{ce}, z_{cb}]$ for a BJT, and $[g_m, z_{gs}, z_{ds}, z_{dg}]$ for an FET. These virtual values are

unique for a specific topology, and they are reversibly transformable from standard two-port matrix parameters¹⁶.

The plots in section 3.2.1, show virtual parameter plots of *measured* transistor data. These plots represent the behavior models of figure 2.17b and c. The fact that their values match very well with ideal elements demonstrates how attractive this novel descriptive method can be.

2.3.3. Conclusions

Standard matrix methods for the representation of linear two-ports have been summarized, being an important group of multi-port matrix parameters.

We propose in a natural way the identification of a new class of network parameters: *virtual circuit parameters*. They are convenient for assessing the validity of simple lumped element approximations to distributed networks. In addition, a convenient set of virtual circuit parameters for BJT's (bipolar junction transistors) and FETs (field effect transistors) has been selected. A set of virtual parameter plots of measured transistor data combines the advantages of an exact *linear* specification with those of simple though adequate transistor models.

This section provided the mathematical tools for transforming measured two-port data into a form that is most convenient for the given application.

¹⁶ Keep in mind that the transistor is assumed to be biased linearly!

Chapter 3

Extraction of transfer parameters

The quantification of parasitics, as discussed in section 1.2.3, is of basic importance for wideband feedback design. Extraction of parameter values from measurements is a robust approach to quantify parasitics.

Low frequency measurements, usually below 1 MHz, are often used to estimate device parasitics. The application of estimated parameters from low frequency measurements are restricted to simplified device models. They may fail, however, for complex devices such as transistors and at 'high' frequencies.

This chapter introduces some new and powerful extraction methods to model devices from measured data, using equivalent circuits that include parasitic interactions. Examples demonstrate the applicability of this work to semiconductor devices such as photo-diodes and transistors. All discussions are focused on *linear* measurements and modeling.

Highlights of this chapter

The highlights of this chapter, which have been explored in this study, are:

- Development of a robust algorithm to fit rational functions with measured transfer functions (magnitude and phase). The iterative method is simple and requires no starting values. Further, various related algorithms are developed for extracting rational functions when transfer information is incomplete. For instance, spectral measurements performed with spectrum analyzers and white noise as stimulus are scalar in nature, so that phase information is lacking.
- Development of a robust algorithm to extract simple transistor models (BJT's and FET's) from measured two-port parameters for synthesis purposes. The method is significantly simpler than four-dimensional brute-force methods, as for instance are implemented in circuit simulators such as Touchstone® [124].
- Discrepancies found between measurements and the widely accepted theory [406] that diffusion capacitance is the dominant factor in limiting the transition frequency of bipolar transistors. The two-port measurements we performed and the measurements specified by semiconductor manufacturers indicate that diffusion capacitance might be a quantity of minor importance in wideband BJT's.
- Development of a simple (linear) bipolar junction transistor model, that is adequate for synthesis purposes and valid for frequencies up to the transition frequency f_T . This is a significant improvement compared to conventional models based on diffusion capacitance concepts. Conventional models are inadequate for frequencies in excess of 10% of f_T .

This chapter provides a theoretical foundation for the rest of this work.

3.1. One-port model extraction methods for linear synthesis

Wideband circuit design may fail when using models, on which the element values are estimated from low frequency measurements.

The introduction [213,214] in 1967 of automated microwave network analyzers has facilitated linear transfer measurement of devices, observed relative to one or more reference planes. The blackbox methods of section 2.2 facilitate the representation of these measurements with matrix parameters for each frequency of interest. These representation methods are eminently suited to wideband circuit design, because various circuit simulators can handle matrix parameters in tabular format as easily as they handle equivalent circuit models.

This approach, although effective for computer assisted circuit analysis, is inconvenient for circuit synthesis¹. The complexity of synthesis forces the use of models that are *as simple as possible*. It requires models with a minimum number of parameters, yet still adequate over the full frequency band of interest. This is crucial, since wideband synthesis methods rely in the main on manual interpretation of device parameters.

This work has resulted in robust algorithms for extraction of device models from measured data, suitable for (linear) synthesis. The modeling algorithms are based on the extraction of rational functions (poles and zeros) and have been proven useful in avoiding superfluous elements² in device models. These rational functions provide polynomial coefficients or poles and zeros that are directly related to the element values of an equivalent circuit model.

This section discusses the methods of extraction in detail.

The discussion that will follow on device modeling and parameter extraction is based on profound experience with practical measurements. This requires more than a single measurement instrument, and therefore we realized in several years a design environment for measurement, data-acquisition, analysis and synthesis. The major part of the SABEL-CAE³ system has been described in many internal reports, and a discussion of all its aspects is beyond the scope of this text. A short overview of its application has been described in [101,109] and further in section 1.4.2.

3.1.1. Elementary extraction methods of analytical transfer functions

The most elementary extraction is the determination of poles and zeros or of polynomial coefficients from transfer functions in tabular form. Let $\mathbf{H}$ and ω be two column vectors, representing the (measured) complex function values and their associated angular frequencies. Curve fitting with an appropriated analytical function will provide usable parameters for modeling purposes. In this section, we focus on (one of) the following target functions:

$$H_1(j\omega) = H_1(s) = a_0 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_n \cdot s^n$$

¹ The purpose of analysis is to find the properties of an extant circuit. Circuit simulations as well as measurements are adequate methods. The purpose of synthesis is to find a circuit that meets a set of pre-defined conditions. This is the reverse problem.

² It is not uncommon to add series inductors to all output terminals of a device model, to represent bond wires. We observed that most of these additions are superfluous in applications below 1 GHz.

³ SABEL-CAE: Synthesis and Analysis of Broadband ELectronics, using Computer Assisted Engineering.

$$H_2(j\omega) = H_2(s) = G_\infty \cdot \frac{a_0 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{m-1} \cdot s^{m-1} + s^m}{b_0 + b_1 \cdot s + b_2 \cdot s^2 + \dots + b_{n-1} \cdot s^{n-1} + s^n}$$

$$H_3(j\omega) = H_3(s) = G_0 \cdot \frac{(1-s/z_1) \cdot (1-s/z_2) \cdot (1-s/z_3) \cdot \dots \cdot (1-s/z_m)}{(1-s/p_1) \cdot (1-s/p_2) \cdot (1-s/p_3) \cdot \dots \cdot (1-s/p_n)}$$

$$H_4(j\omega) = H_4(s) = G_\infty \cdot \frac{a_0 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{m-1} \cdot s^{m-1} + s^m}{b_0 + b_1 \cdot s + b_2 \cdot s^2 + \dots + b_{n-1} \cdot s^{n-1} + s^n} \cdot e^{-s\tau}$$

For first order estimations, a manual curve fit is adequate. It relies on the estimation of asymptotic lines of transfer functions, using the Bode plot. The corner frequency of these asymptotic lines equals the corner frequencies of a pole or a zero.

The higher the order⁴ of approximation, the more the extraction relies on automated curve fitting. Some examples are:

- Automated polynomial fits (Taylor expansions) to $H_1(j\omega)$ require (non-iterative) linear extraction methods as described in appendix C.
- Automated rational function fits to $H_2(j\omega)$ require iterative extraction methods as described in appendix D. This is a new (linear) iteration method that does not require starting values. This has made the algorithm robust and simple.
- Automated pole-zero fits to $H_3(j\omega)$ are similar to rational function fits to $H_2(j\omega)$, and use an additional root finding algorithm for polynomials.
- Automated delayed pole-zero fits, to $H_4(j\omega)$, require iterative *non-linear* extraction methods. Well-known iteration algorithms, such as Levenberg-Marquardt [410,413] or Gauss-Newton [413], are applicable. The development of simpler and robust algorithms, using linear methods, is of interest for further investigation. Alternatively, a constrained rational phase fit is often applicable, as discussed in the succeeding paragraph.

Optimal curve fitting requires the minimization of some residual error term in a least squares sense. Various algorithms found in the literature minimize some *absolute* error, such as $\epsilon_1(j\omega) = |H(j\omega) - h(j\omega)|$ or $\epsilon_2(j\omega) = (H(j\omega) - h(j\omega))^2$. One sometimes constructs rational function approximations from a power series expansion of $h(j\omega)$. This is called Padé approximation [413]. The Remez algorithm [413] produces the best Chebyshev approximation to a given frequency response with a fixed number of filter coefficients.

We have found it more useful to minimize the *relative* error $\epsilon(j\omega) = H(j\omega)/h(j\omega) - 1$ in a least squares sense over all angular frequencies ω . In this expression $h(j\omega)$ is the data table and $H(j\omega)$ is the function to be fitted. Our algorithms facilitate weighing of this relative error for enhancing curve fitting in specific frequency intervals.

The linear extraction algorithms in this section, as well as many other algorithms that will be discussed in this book, rely on methods that solve over-determined sets of linear equations. These methods use an over-determined matrix division, as described in detail in appendix B.

Constrained curve fits

⁴ The approximation order is the number of roots in a polynomial fit, or the highest number of poles or zeros in a rational function fit.

Several algorithms are applicable when the magnitude as well as the phase of the transfer function is measured. The acquisition of this information is feasible with vector network analyzers, which are commercially available.

In some practical situations, however, transfer information is limited to either magnitude or phase information. This work has resulted in two additional automated extraction algorithms, as is described in appendix *E* and *F*.

- A rational magnitude fit extracts $H(j\omega)$ from $|H(j\omega)|$ in tabular format. The network is assumed to be a minimum phase network, which means that all poles and zeros have negative real part.
- A rational delay fit extracts $H(j\omega)$ as a pseudo delay transfer function. The network is assumed to be an all-pass network, which means that $|H(j\omega)| \equiv 1$

A constrained magnitude fit is required, when the transfer is measured with a spectrum analyzer using a tracking generator or white noise generator as source. Other examples are scalar network analyzer measurements.

3.1.2. Extraction methods for equivalent circuit elements

Elementary transfer function curve fits generate impedance or admittance functions without regard to physical implementation. This simplifies determination of a suitable order of approximation. Pole-zero extraction methods are generally applicable to arbitrary *lumped* element models. Polynomial curve fits are of limited applicability, and are usually restricted to the extraction of dominant properties of the transfer function. Note that a *polynomial* fit of $H(j\omega)$ maps the transfer function to ∞ for infinite frequencies, and a fit of $1/H(j\omega)$ maps the function to zero. Polynomial curve fits are nevertheless preferred when applicable, since they are non-iterative methods.

Polynomial fits

When Z is a lumped impedance function, then polynomial methods are of practical applicability to the following combinations of lumped impedances:

<i>series R and L</i>	$Z = R + j\omega \cdot L$
<i>series R and C</i>	$j\omega \cdot Z = 1/C + j\omega \cdot R$
<i>series R, C and L</i>	$j\omega \cdot Z = 1/C + j\omega \cdot R + (j\omega)^2 \cdot L$
<i>parallel R and C</i>	$1/Z = 1/R + j\omega \cdot C$
<i>parallel R and L</i>	$j\omega/Z = 1/L + j\omega \cdot R$
<i>parallel R, C and L</i>	$j\omega/Z = 1/L + j\omega \cdot R + (j\omega)^2 \cdot C$

Examples are discussed in subsection 3.1.3 concerning impedance extraction of lasers and photo diodes. These examples are special situations of Taylor series. These (infinite) series are applicable when their interval of convergence spans the frequency band of interest. Unfortunately, this applies only for a limited⁵ set of transfer functions.

⁵ An example that illustrates when it is useless to evaluate a higher order Taylor expansion is a shunt of series impedances with R and C . The most striking aspects of overall impedance are observed near the corner

Examples of the successful use of a Taylor expansion are simple distributed models with pure delay. This is because an $\exp(x)$ function is convergent for all values x . This is demonstrated in section 3.2 in the context of transistor model extraction.

Polynomial methods are occasionally of practical applicability to functions as:

$$\begin{aligned} e^{-j\omega\tau} &= 1 - j\omega\tau + 1/2 \cdot (j\omega\tau)^2 - 1/6 \cdot (j\omega\tau)^3 + \dots \pm 1/n! \cdot (j\omega\tau)^n + \dots \\ e^{-j\omega\tau_d} \cdot (1 - j\omega\tau_0) &= 1 - j\omega(\tau_0 + \tau_d) + 1/2 \cdot (j\omega)^2 \cdot (\tau_d^2 + 2 \cdot \tau_0 \tau_d) + \dots \end{aligned}$$

Rational function fits

Pole-zero extraction methods are more generally applicable, although they require an iterative approach. A rational function fit, as described in appendix D, extracts the coefficients of the numerator and denominator polynomials and is relatively simple.

On the other hand, the arithmetic precision of the computer is a serious limitation in the application of the method. Using double precision (15 decimals), the ratio between the largest and smallest coefficients is limited to 10^{15} . Adequate frequency scaling⁶ will balance the polynomial coefficients to minimize this disadvantage. As a result, an extraction of rational functions with 6th-order polynomials is limited to a frequency interval not exceeding $(f_{\max}/f_{\min}) \approx 10^5 = \sqrt[3]{(10^{15})}$.

In exceptional situations, this interval limit is unacceptable. Alternative methods that focus directly on poles and zeros, instead of polynomial coefficients, are potentially applicable over wider frequency intervals. They are more complicated than the (linear) rational function fits and requires the use of non-linear iteration methods. Examples of these methods are Levenberg-Marquardt and Gauss-Newton, based on partial derivatives.

In summary, our linear rational extraction methods are often adequate for the extraction of models for synthesis purposes. In many applications, the non-iterative polynomial method is even adequate.

3.1.3. Extraction of adequate photo-diode impedance-models

When designing lightwave receivers using feedback loops, the impedance of the photo-diode affects the loopgain transfer function. Analysis and synthesis of loop stability as well as receiver bandwidth rely on adequate models of the source impedance of this input device. This illustrates that source impedance is an important design parameter.

A first-order approximation of this source is a controlled current source, proportionally controlled by the light intensity. Its responsivity is frequency dependent. Since this responsivity does not affect the loop stability of the lightwave receiver and may be chosen to be flat over a wide frequency interval, this design parameter is of secondary importance.

frequencies $\omega=1/RC$ of the individual shunts. Each shunt admittance equals $Y=j\omega C/(1+j\omega RC)$ and generates a Taylor expansion, similar to $x/(1+x)=x \cdot (1-x+x^2-x^3+x^4\dots)$. The interval of convergence begins or ends when $|x|=1$ or when $\omega<1/RC$ or $\omega>1/RC$. As a result, in this example the Taylor expansion is unable to model the most striking aspects of the transfer function and yields poor accuracy results.

⁶ An adequate choice of frequency ω_0 balances the leading and trailing coefficients of a polynomial $p(s)$. It equals $q_3=1$ in: $p(s) = c_0 + c_1 \cdot s + c_2 \cdot s^2 + c_3 \cdot s^3 = c_0 \cdot \{ 1 + q_1 \cdot (s/\omega_0) + q_2 \cdot (s/\omega_0)^2 + q_3 \cdot (s/\omega_0)^3 \}$. The magnitude of one of the center coefficients will probably dominate the other coefficients.

Impedance measurements will demonstrate whether a simple impedance approximation is adequate or that higher order models are necessary. For this, it is required that the test geometry simulates the situation in which the device is connected to the circuit. For instance, a home-made PCB (printed circuit board) test-fixture shown in figure 3.1 is appropriate when the photo-diode is mounted on the edge of a PCB with microstrip layout. The photographs in figure 3.4 and 3.5 show an alternative construction in case the DUT has special mounting requirements.

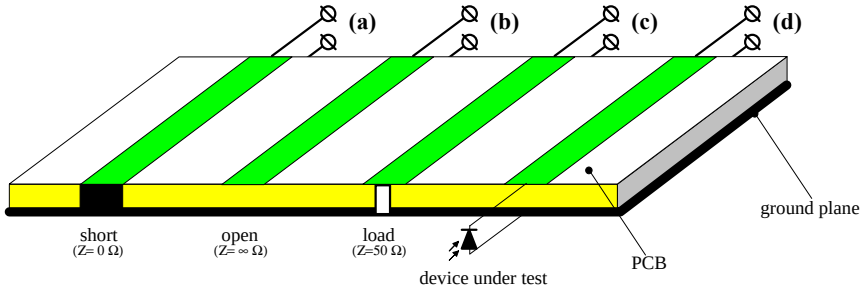


Fig 3.1 Example of a test-fixture with microstrip layout to measure the impedance of photo-diodes. It is an epoxy PCB (printed circuit board) with 1.5 mm thickness, on which the device under test is mounted on the edge. Its unknown impedance is compared to three well-known impedances using a network analyzer: a Short (0Ω) an Open ($\infty \Omega$) and a Load (e.g. a 50Ω SMD-resistor) mounted on the edge of the PCB test-fixture

Figure 3.1 shows a simple test fixture to measure the photo-diode impedance observed relative to a reference plane that coincides with the edge of the PCB. A vector network analyzer is subsequently connected to the ports *a*, *b*, *c* and *d*, and the device impedance is extracted from all four reflection measurements and the three known impedance values. This is a standard feature on network analyzers, when calibrated with a "Short", "Open" and "Load" on port *a*, *b* and *c*.

Figure 3.2 shows the measured impedance, observed relative to the edge of the PCB test-fixture. To demonstrate the use of the impedance extraction methods in subsection 3.1.2, we applied various curve fits to this measured impedance. The following fits are shown in figure 3.2:

- A simple first order polynomial fit (Z_1), fitted up to 500 MHz, provides a first order approximation that is adequate up to 1 GHz.
- A second order polynomial fit (Z_2), fitted up to 2.2 GHz provides a model that is applicable up to 2.2 GHz.
- A third order rational function fit, fitted over the full measured frequency band provides reliable modeling up to 4 GHz.

The measured impedance illustrates a decreasing impedance with increasing frequency. This is indicative of dominant capacitive behavior.

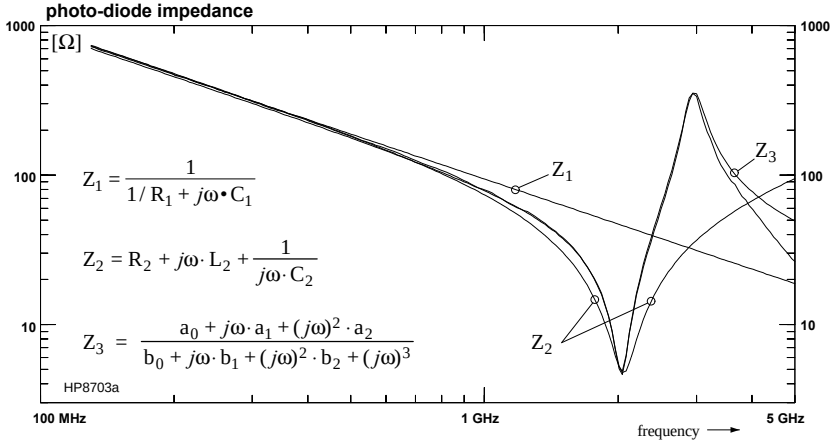


Fig 3.2 Measured impedance Z of a photo-diode, observed relative to the edge of the PCB test-fixture, as is shown in figure 3.1. This function is fitted with various curves using polynomial and rational functions.

Figure 3.3 shows some simple models of photo-diodes. The controlled current sources represent the responsivity of the photo-diode, which is frequency dependent. Nevertheless, this transfer function is commonly flat over a wide frequency band. Modeling its transfer requires additional measurements, which are not considered in this subsection.

In this example, the extracted resistance values of the shunt RC-model as well as series RC-model are not significant. This is because the influence of the (very high) shunt resistor is noticeable far below the measured frequency interval, and the influence of the (low) series resistance is negligible comparing to higher order effects. As a result, the only relevant elements in the first order models are the extracted capacitors.

On the other hand, the extracted series resistor of the series RCL model is meaningful. Variation of that value will significantly reduce the accuracy of the model.

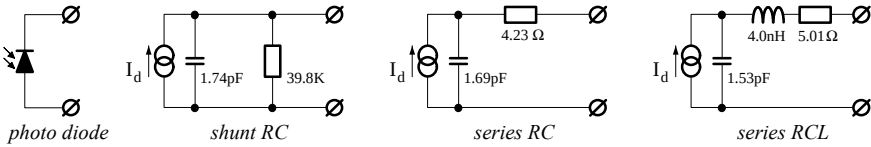


Fig 3.16 Examples of simple (impedance) models of photo-diodes, extracted from the measurements in figure 3.2.

Note that the extracted capacitor values in figure 3.3 are slightly different (10%). This demonstrates that specification of diode capacitance is model-dependent. As a result, high frequency extrapolation, based on low frequency capacitance measurements, is simply not possible. Most (higher order) curve fitting is not associated with a specific circuit topology. It illustrates rather that rational function fits serve as an intermediate step between measured data and an extracted circuit model. This approach facilitates an assessment of the value of higher order approximations, while ignoring the circuit topology required. Modeling of marginal (higher order) side effects is avoided by using

rational function fitting. As a result, our rational function approach yields the simplest equivalent circuit model, that is adequate for circuit analysis and synthesis, within the frequency band of interest.

3.1.4. Conclusions

This study resulted in iteration techniques for fitting rational functions with measured data in tabular form. The algorithms are linear and require no starting values. These properties have made the algorithm robust and simple.

One-port extraction methods have been developed, for extracting device models from measured data. The methods rely on polynomial fits or on rational function fits to provide poles and zeros or polynomial coefficients. These coefficients are directly related to the element values of an equivalent circuit model.

The advantage of our extraction approach is that it tackles the complex modeling in isolation from the associated circuit topology. It facilitates an easy assessment as to the need for higher order models, limiting the extraction to the simplest model within the frequency band of interest. This advantage is crucial when extracting models for synthesis purposes.

We have demonstrated the validity of our method by extracting lumped element models of the output impedance of photo diodes. To measure this impedance, observed relative to well-defined reference planes, simple test-fixtures were developed for simulating the situation in which these devices under test are connected to circuits.

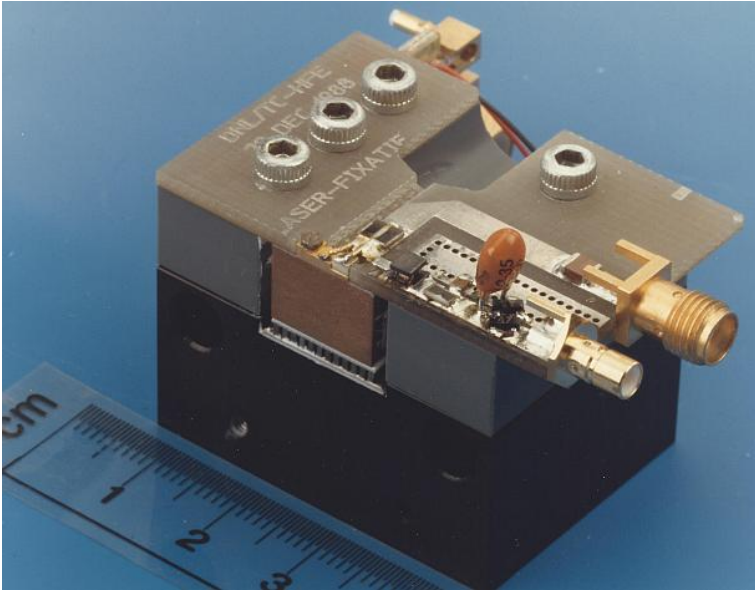


Fig 3.4 Laser chip mounted in a module. Measurement of the chip impedance requires the positioning of a reference plane at the chip location.

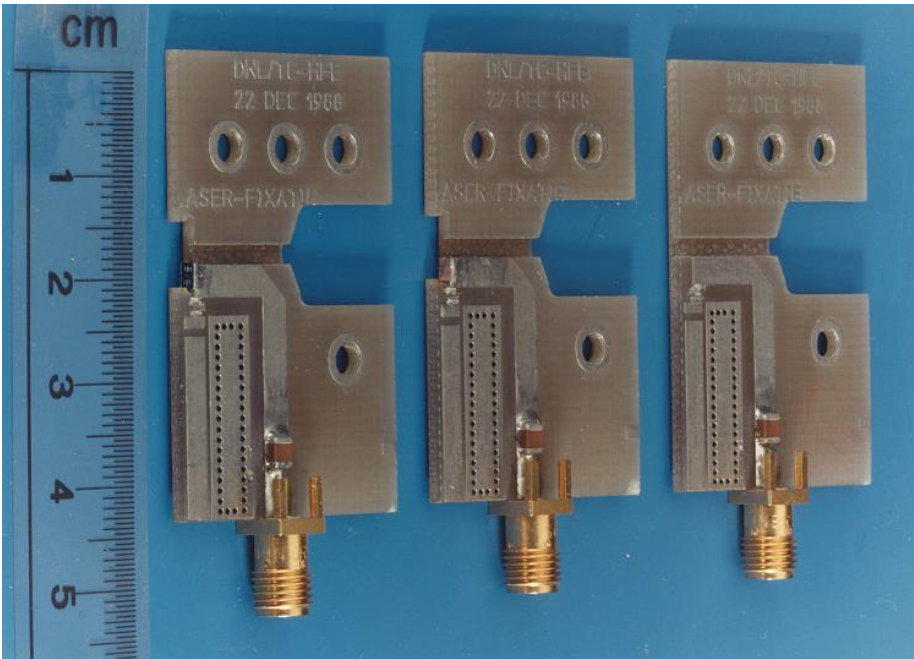


Fig 3.5 Calibration setup for the laser module of figure 3.4. Three duplicate PCB's, on which the laser chip is replaced by an Open, a Short and a Load respectively, to position the reference plane at the chip location.



Fig 3.6 Calibration setup for two-port transfer measurements on transistors. The (SMD) transistor is mounted in the center of the fixture. The symmetry of the setup facilitates the positioning of reference planes in the center of the fixture using a Load, a Short and an Open

3.2. Two-port extraction methods for transistor models

Two-port extraction methods are significantly more complicated than one-port extraction methods. Moreover, the use of adequate *methods* as well as adequate *models* is far more critical. When the method fails, a solution will never be found. When the model fails, the solution is inadequate. Both possibilities will be discussed.

A commonly used extraction method. There are various methods to extract equivalent circuits. For instance, the commercially available software package Touchstone® [124] uses a four-dimensional iterative curve fit to fit (optimize) all relevant circuit elements of the model simultaneously with all four s-parameters of the two-port.

This brute-force method is applicable to arbitrary circuit topologies and therefore indispensable when improving sophisticated circuit models. On the other hand, the computational effort associated with the approach is very high, and the iteration may suffer from convergence problems when poor starting values are supplied.

The optimizer of Touchstone® [124] uses s-parameters as an optimization target. Usually, the relation between these parameters and the equivalent circuit elements is complex. As a result, the use of s-parameters as an optimization target can conceal the real cause when an iteration fails due to an improperly chosen circuit topology.

Wideband circuit synthesis requires equivalent models that are as simple as possible in order to be tractable for the synthesis. The use of iterative brute force methods is unnecessarily complicated when evaluating synthesis models.

This section 3.2 proposes novel extraction methods for transistor synthesis models, that are significantly simpler and require no starting values.

A commonly used transistor model. Adequate small-signal models of the active devices are vital for designing wideband feedback amplifiers. The well-known hybrid- Π model [304,406] is commonly considered to be particularly suitable for describing the small-signal behavior of BJT's (bipolar junction transistors) and FET's (field effect transistors). This BJT-model is based on diffusion capacitance to model its frequency response.

When designing wideband lightwave receivers, we have observed a disagreement between transfer measurements and circuit analysis when applying BJT models based on diffusion capacitance. This disagreement holds particularly for frequencies above 10% of the transition frequency f_T . The same disagreement was observed when applying brute force extraction methods to this conventional BJT model on measured two-port parameters. The hybrid- Π model was observed to be adequate for FET's.

To increase the predictive performance of transistor models, a considerable effort has been devoted to developing accurate and complex device models with many parameters [302,303,305]. Moreover, each minor parasitic effect, such as bond wire inductance or stray capacitance in the package, has often been taken into account in these models. This approach, while effective for analysis, is unsuitable for the purposes of synthesis.

Synthesis requires models that are made as simple as possible, without unacceptable loss in predictive performance. Models based on diffusion-capacitance are commonly used for this purpose [406], however, they become inadequate for frequencies in excess of 10% of f_T . This section discusses in an improved synthesis model for BJT's, and a well-known adequate synthesis model for FET's.

3.2.1. Virtual circuit parameters as device representation method

Our wideband transistor models⁷ rely on full two-port measurements to validate the models for the frequency interval of interest. This requires a reliable measurement method as well as an adequate way to represent the measured results. We will discuss them both.

Measurement method

Reliable measurements on transistors require the test geometry to simulate the situation in which the device will be connected to circuits. When using automated network analyzers, this situation is easily simulated. For instance, a home-made PCB test fixture shown in figure 3.7 is appropriate for the required two-port measurements when the transistor is mounted on the surface of the PCB. A photograph of this setup is shown in figure 3.6.

The calibration of the network analyzer requires an "Open", "Short" and "Load", similar to the construction of figure 3.1. They have been realized by halving a duplicate test-fixture, and mounting the reference impedances on the edge of the PCB, to position the reference plane in the center of the transistor fixture. A calibrating "Thru" has been realized using another duplicate fixture, with a through-connection in the center.

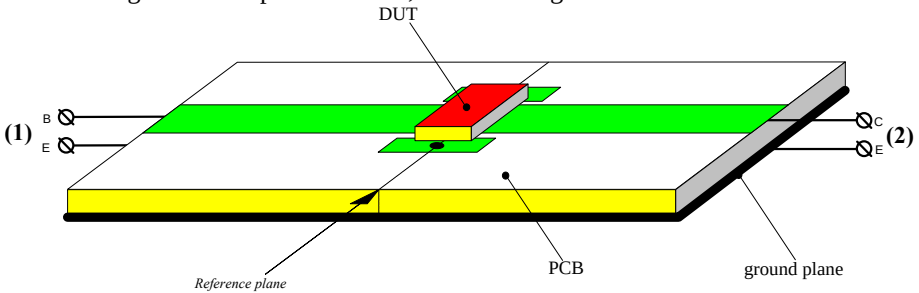


Fig 3.16 Example of a test-fixture with microstrip layout to measure the two-port parameters of transistors. It is an epoxy PCB (printed circuit board) with 1.5 mm thickness, on which the DUT (device under test) is mounted on the surface of the PCB. Via holes connect the emitter contact with the ground-plane on the bottom side.

BJT example

Commercially available network analyzers determine two-port parameters in s-parameter format. We prefer virtual circuit parameters as a representation method (see section 2.3.2) to simplify interpretation and model extraction. The virtual circuit parameters are converted from s-parameters, using y-parameters as an intermediate step. Figure 3.8 shows⁸ measured⁹ virtual BJT-parameters $[\alpha_b, z_e, z_{ce}, z_{cb}]$ of a bipolar junction transistor. The magnitude of the current gain α_e in figure 3.8a illustrates that the transition frequency f_T is approximately 3.5 GHz. This is estimated from the frequency where the current gain parameter α_e has been reduced to $|\alpha_e|=1$. See section 2.3.2 for the definition of the virtual circuit parameter α_e .

⁷ All our transistor modelling effort was restricted to (linear) small signal models

⁸ The measured curves are overlaid with curves predicted by a BJT-model, as discussed in subsection 3.2.2.

⁹ In fact, the s-parameters of the BJT are measured, and the virtual BJT parameters are extracted from them.

Figure 3.8b shows the phase of α_b and α_e converted from measured s-parameter data. These plots illustrates that our BJT-model provides an excellent fit with the measured phase over the full frequency band. The associated phase-delay and group-delay¹⁰ (not shown in this plot) are almost frequency independent. Figure 3.8c and 3.8d show the three impedance values of the virtual BJT-parameters. These curves provide an excellent fit for z_e and z_{cb} and a fair fit for z_{ce} . This demonstrates the validity of our BJT model.

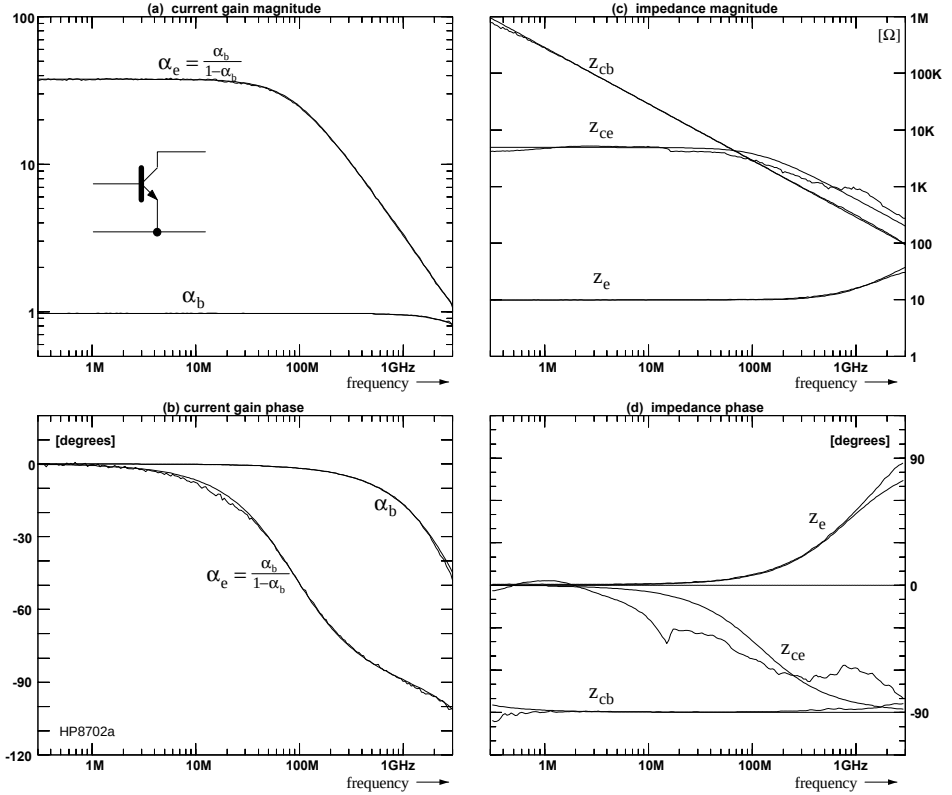


Fig 3.16 Virtual circuit-parameter plots to represent the measured two-port parameters of a bipolar junction transistor. These parameters are defined in section 2.3.2. The dotted lines are converted from measured s-parameter data (201 frequency points), and the solid lines are predicted by the BJT model proposed in figure 3.10. That model is more adequate than the well-known hybrid P-model, as discussed in section 3.3.2.

FET example

Figure 3.9 shows measured virtual FET-parameters [g_m , z_{gs} , z_{ds} , z_{dg}] of a field effect transistor (CF910: dual-gate GaAs-MesFET), superimposed on predicted virtual parameters using a hybrid-II model. These parameters are slightly different from virtual BJT parameters, however, a similar approach applies.

¹⁰ The phase delay of phase $\phi(\omega)$ is defined as: $\tau_{dp} = -\frac{\phi}{\omega}$, and the group delay is defined as $\tau_{dg} = -\frac{d\phi}{d\omega}$

The curves in figure 3.9a and 3.9b demonstrate an excellent fit for the transconductance g_m over the full frequency band. The associated phase-delay and group delay (not shown in this plot) are almost frequency independent.

A similar resemblance applies for the reconstructed impedances z_{dg} and z_{gs} in figure 3.9c and 3.9d, although the poor resemblance of their phase suggests the contrary. This is a seeming contradiction since z_{gs} and z_{dg} represent very weak effects, especially at low frequencies, because their shunt impedance is very high. Measurement of these weak effects is beyond the accuracy limits of the used measurement setup.

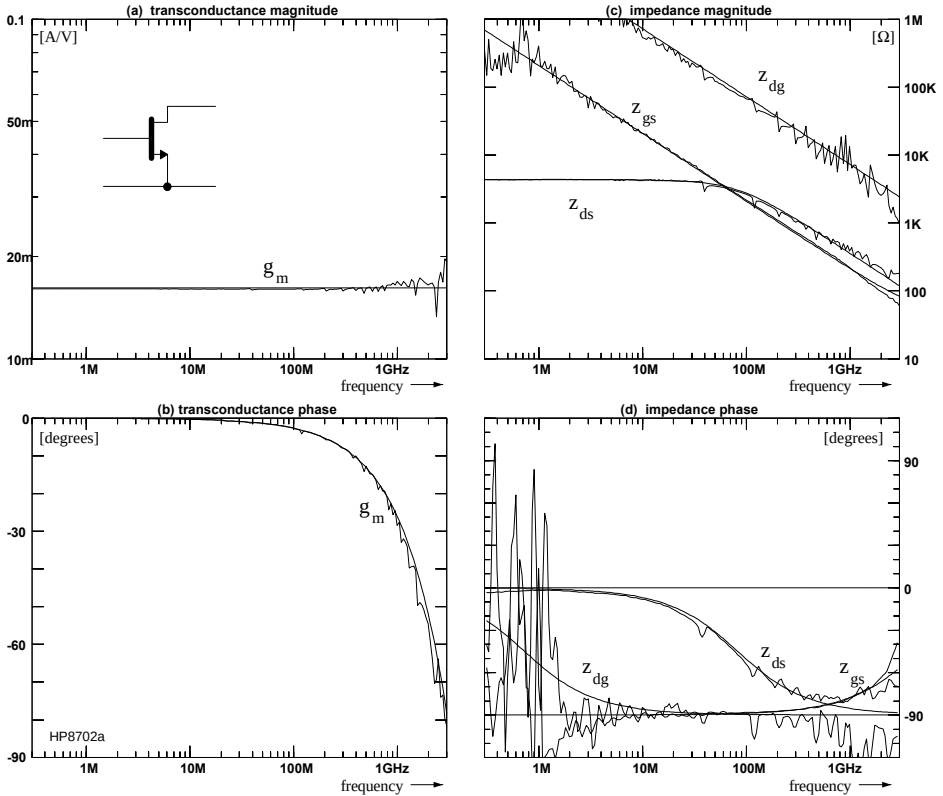


Fig 3.16 Virtual circuit-parameter plots to represent the measured two-port parameters of a field effect transistor. These parameters are defined in section 2.3.2. The dotted lines are converted from measured s -parameter data (201 frequency points), and the solid lines are predicted by the FET model shown in figure 3.11.

3.2.2. Extraction of transistor parameters using Taylor series expansion

Brute force methods use a four-dimensional iteration for fitting all two-port parameters simultaneously with the model. We developed an alternative method that use four *individual* one-dimensional fits. It is applicable to transistors when using adequate virtual circuit parameters as two-port representation.

Our approach is significantly simpler and normally requires no iteration at all. It consists of the following steps:

- First, transform the measured two-port parameters to a virtual circuit-parameter format that is closely related to the circuit topology of interest. For transistors, these parameters are defined in section 2.3.2 and examples are shown in figure 3.8 and 3.9.
- Fit each virtual circuit parameter individually with a simple rational or polynomial function. This approach provides (polynomial) coefficients that are closely related to the circuit elements of an intrinsic circuit model.
- If this intrinsic circuit model is inadequate for the frequency interval of interest, extend the model with additional elements. Use the element values of the intrinsic model as starting values in a four-dimensional iteration.

In our experience, the first two steps are often adequate for transistor model extraction. Furthermore, it is observed that simple polynomials (Taylor series, expanded to one or two power terms) are often adequate for fitting measured virtual circuit parameters. This provides the means to define the most significant transistor parameters in an unambiguous way. Extraction of these parameters is performed in a similar way.

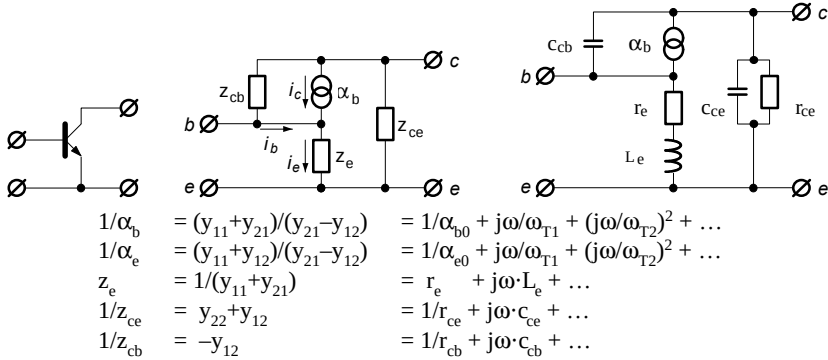


Fig 3.16 Simple BJT transistor model of which the parameters can easily be extracted from Taylor expansions of measured two-port parameters transformed to virtual BJT parameters. We define:

$$\omega_T = \omega_{T1}/\alpha_{b0} = 2\pi \cdot f_T, \quad \omega_e = r_e/L_e \quad \text{and} \quad c_e = 1/(r_e \cdot \omega_T). \quad \text{Note that } 1/\alpha_b = 1/\alpha_e + 1.$$

The equations in figure 3.10 define the most significant parameters of bipolar junction transistors as the coefficients of a Taylor expansion of the associated virtual circuit parameters. They are directly applicable in equivalent circuit models. Furthermore, they illustrate the advantages of using virtual circuit-parameter plots as a device representation method. The most significant transistor parameters are directly obtainable from graphical interpretation of the logarithmic plots using their asymptotic slopes.

Figure 3.11 provides similar definitions for field effect transistors. Note that the expansion for z_{gs} is, strictly speaking, a MacLaurin expansion.

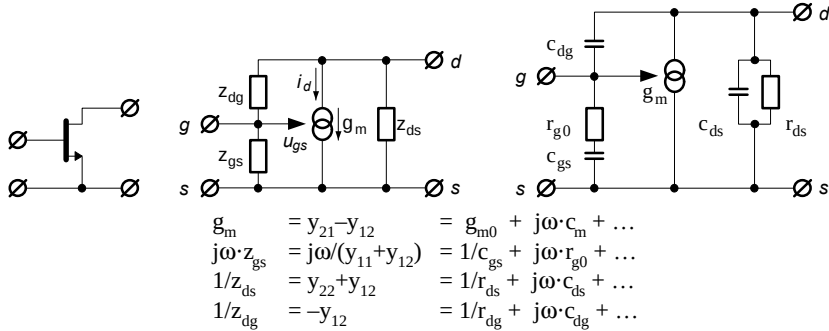


Fig 3.16 Simple FET transistor model of which the parameters can easily be extracted from Taylor expansions of measured two-port parameters transformed to virtual FET parameters.

We define: $\omega_T = g_m / (C_{gs} + C_{dg})$

In figure 3.8 and 3.9 we demonstrated that simple Taylor series expansion of transistor parameters are applicable over a very wide frequency band. Furthermore, we have observed this for a variety of wide band transistors. An example is shown in figure 3.12, under varying the bias conditions.

BFR92A NPN bipolar transistor

I_c	2	5	10	20	2	5	10	20	mA	
U_{ce}	5	5	5	5	10	10	10	10	V	
α_{e0}	68.949	73.153	78.993	83.133	75.435	81.664	90.139	103.066	---	see section 3.2.3
f_T	2.972	4.895	6.245	6.865	2.937	4.874	6.265	6.910	GHz	
$\tau_{\alpha 0}$	16.085	42.075	65.907	100.000	18.477	37.722	100.000	100.000	%	
$\tau_{\alpha 1}$	83.915	57.925	34.093	0.000	81.523	62.278	0.000	0.000	%	
r_e	12.364	5.883	3.464	2.220	12.884	6.198	3.766	2.529	Ω	
L_e	1.426	1.531	1.526	1.530	1.434	1.549	1.552	1.563	nH	
r_{ce}	34.655	20.322	9.877	7.246	64.433	64.027	10.727	8.373	k Ω	
c_{ce}	0.230	0.240	0.243	0.243	0.232	0.234	0.218	0.228	pF	
c_{cb}	0.424	0.421	0.418	0.424	0.392	0.394	0.397	0.400	pF	
f_T/α_{e0}	43.100	66.918	79.057	82.576	38.930	59.685	69.501	67.041	MHz	
f_{T1}	2.929	4.829	6.167	6.783	2.898	4.815	6.196	6.843	GHz	$= \omega_{T1}/2\pi$ $= \omega_{T2}/2\pi$
f_{T2}	7.671	8.435	9.336	9.334	7.124	8.757	8.800	9.159	GHz	
f_e	1.380	0.612	0.361	0.231	1.430	0.637	0.386	0.257	GHz	
c_e	4.332	5.527	7.357	10.444	4.206	5.268	6.747	9.108	pF	

Fig 3.12 Example of extracted transistor parameters, using the s -parameters as specified by the manufacturer (Philips). Note that all values r_e are approximately 1.3W higher then is expected from the bias current I_c . This is due to emitter series resistance, which is roughly independent of the bias current.

3.2.3. Extraction of parameters with delay using Taylor series expansion

The validity of the transistor models in figure 3.10 and 3.11 can be improved when modeling the gain transfer functions as the cascade of a low-pass transfer function and a delay function. The modified transfer functions becomes:

$$BJT: \quad \alpha_b = \frac{\alpha_{b0}}{(1+j\omega\tau_{\alpha1} + \dots)} \cdot e^{-j\omega\tau_{\alpha0}}$$

$$FET: \quad g_m = \frac{g_{m0}}{(1+j\omega\tau_{g1} + \dots)} \cdot e^{-j\omega\tau_{g0}}$$

One way to extract these modified gain transfer functions is by using an iterative curve fit to the proposed function. A simpler (non-iterative) way is evaluating a Taylor expansion and using the extracted coefficients for calculating the modified delay coefficients. Delay is represented by infinite Taylor series, however, it can be reconstructed when the first terms of the expansion are known. For instance, when the expansion of α_b is restricted to the first three terms, it provides the coefficients (α_{b0} , ω_{T1} , ω_{T2}) as defined in subsection 3.2.2. These parameters are related as follows to the modified delay coefficients ($\tau_{\alpha0}$, $\tau_{\alpha1}$):

$$BJT: \quad \frac{\alpha_{b0}}{\omega_{T1}} = \tau_{\alpha0} + \tau_{\alpha1} \quad \frac{\alpha_{b0}}{(\omega_{T2})^2} = \tau_{\alpha0} \cdot \tau_{\alpha1} + \frac{1}{2} \cdot (\tau_{\alpha0})^2$$

Modeling with additional delay may improve the gain transfer function for frequencies near the transition frequency ω_T . This is illustrated in figure 3.13 for the current transport factor α_b of a 1 GHz transistor. In these plots, the delay ($\tau_{\alpha0}$) is varied while the (reciprocal) transition frequency $1/\omega_T = \alpha_{b0}/\omega_{T1} = (\tau_{\alpha0} + \tau_{\alpha1})$ is kept constant. Three distinct curves are plotted in figure 3.13:

- A: for $\tau_{\alpha0}=0$ and $\tau_{\alpha1}=1/\omega_T$ (low pass transfer)
- B: for $\tau_{\alpha0}=0.5/\omega_T$ and $\tau_{\alpha1}=0.5/\omega_T$ (mixed delay and low-pass)
- C: for $\tau_{\alpha0}=1/\omega_T$ and $\tau_{\alpha1}=0$ (delayed transfer)

All curves A, B and C have equal magnitude response and equal phase delay for frequencies below 10% of the transition frequency. The more phase delay is originating from pure delay, the less frequency dependency is observed in magnitude response and in phase delay.

The current gain plots in figure 3.8a and 3.8b show that the current transport factor α_b of a BJT differs from a simple first order low-pass response. A second order Taylor expansion provides accurate modeling results ($f_{T1}=3.26$ GHz, $f_{T2}=6.86$ GHz) within the measured frequency interval. The same applies for a first order Taylor expansion with additional delay. In this example the highest measurement frequency is too low to assess whether models with delayed gain are more appropriate or not.

When converting the second order Taylor expansion to a first order variant we observed that 27% of the total phase delay in α_b originated from pure delay while the remaining 73% originated from the phase shift in the low-pass transfer function. These ratios are bias and device dependent, as is shown in figure 3.12. Investigation of this ratio for various transistors, has shown that delay aspects dominate for high bias currents.

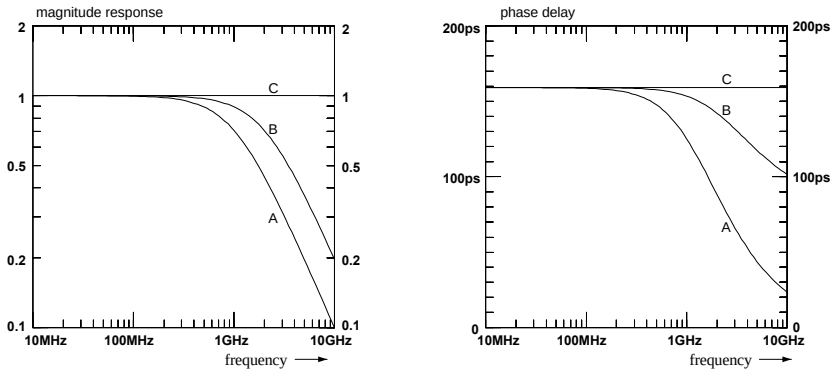


Fig 3.16 Simulated current transport factor α_b of different bipolar junction transistors with equal transition frequency (1 GHz). Curve A illustrates when α_b is modeled by a first order low-pass transfer. Curve C result from a model with pure delay, while curve B is a mix from low-pass transfer and delay.

3.2.4. Conclusions

An attractive graphic method has been discussed for representing measured two-port parameters of transistors. We illustrated the advantages of using these virtual circuit-parameter plots as a device representation method. Most significant transistor parameters are directly obtainable from graphic interpretation of these logarithmic plots using their asymptotic slopes.

Transistor models for (linear) synthesis must be as simple as possible. Over-simplified as well as unnecessary complicated models are inadequate to the job and therefore they must be based on two-port measurements within the frequency band of interest.

A new extraction method has been developed and demonstrated for modeling transistors (BJT's and FET's). Our approach, using virtual circuit parameters as linear two-port description, significantly simplifies two-port modeling. It transforms an all-in-one four-dimensional curve fitting into four simple one-dimensional curve fitting procedures.

In many situations, we have observed that our approach requires no iterations at all, since polynomial fits (Taylor series) are adequate. This observation provides the means for defining the most significant transistor parameters in an unambiguous way.

On occasion, our one-dimensional approach is unable to extract adequate transistor models. In these situations, our approach is a suitable intermediate step for providing excellent starting values to brute-force four-dimensional curve fits.

3.3. Discussion on commonly used transistor models

The two-port extraction methods of section 3.2.2 and 3.2.3 resulted in a non-conventional model for bipolar junction transistors. Commonly used BJT models are extended with base resistance and rely on diffusion capacitance in stead of delay. This section discusses these differences.

3.3.1. Discussion on base resistance versus emitter inductance

The virtual circuit parameter z_e for bipolar junction transistors, as defined in section 2.3.2, is often inductive in nature. One way to model this effect is a series emitter inductor (bonding wires), as illustrated in figure 3.14a. Another way is modeling an inductive z_e parameter with a series base resistor, as illustrated in figure 3.14b.

Both models are associated with equal α_e , z_{ce} and z_{cb} parameter values, and similar values for the z_e parameter: A first order Taylor approximation for z_e yields:

$$\text{model in figure 3.14a: } z_e \approx r_e + j\omega \cdot L_e$$

$$\text{model in figure 3.14b: } z_e \approx (r_{e0} + r_{b0}/\alpha_{e0}) + j\omega \cdot (r_{b0}/\omega_T)$$

In practical situations, for instance the transistor example in figure 3.8, the differences are small. Linear two-port transfer measurements alone are usually inadequate to distinct whether inductive z_e effects are to be assigned to emitter inductance, to base resistance or to a mix of both. We prefer modeling with emitter inductance to simplify the overall circuit analysis.

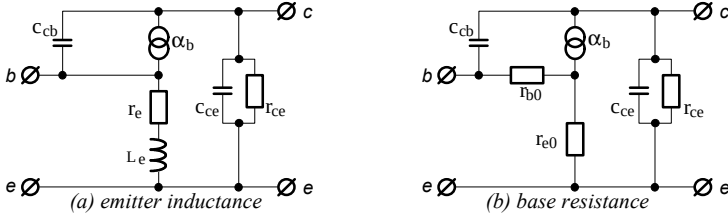


Fig 3.16 Example of two different equivalent T-models for bipolar junction transistors. Emitter inductance (a) as well as base resistance (b) is capable of modeling inductive effects in the virtual circuit parameter z_e (defined in section 2.3.2) of bipolar junction transistors. Both T-models have equal values for the parameter α_e , z_{ce} and z_{cb} , and comparative values for z_e .

When BJT-models with base resistance are preferred, the model in figure 3.10 is to be considered as an intermediate step. This is because its parameter values r_e and L_e are extracted in a structured and unambiguous way. The emitter inductance variant is subsequently transformed into a base resistance variant using the transformation:

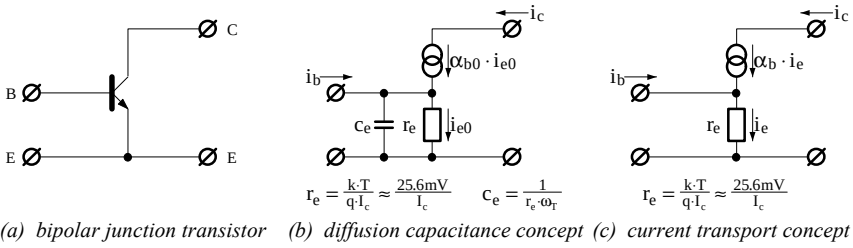
$$r_{b0} = L_e \cdot \omega_T$$

$$r_{e0} = r_e - L_e \cdot \omega_T / \alpha_{e0}$$

3.3.2. Discussion on current transport function vs. diffusion capacitance

The overall current gain of a bipolar junction transistor is low-pass in nature (injected in the input using a current source and sensed at the output when shorted). Our transistor model of figure 3.10 uses a controlled current source α_b to model this effect with frequency dependent current transport factor. This is stressed in figure 3.15c. Diffusion capacitance, as shown in figure 3.15b, is another way to model a low-pass overall current gain. Nevertheless, the overall input impedance of both models are different,

which means that at least one of these models is inadequate. This subsection discusses both concepts.



definition	diffusion capacitance concept $\omega_T, r_e, \alpha_{b0}$		current transport concept $\omega_T, r_e, \alpha_{b0}$		
	relation	transfer	relation	pure low-pass	pure delay
$\alpha_b = \frac{i_c}{i_e}$	$\alpha_b = \frac{\alpha_e}{1+\alpha_e}$	$= \frac{\alpha_{b0}}{1+j\omega/\omega_T}$	$\alpha_b = \frac{\alpha_e}{1+\alpha_e}$	$= \frac{\alpha_{b0}}{1+j\omega/\omega_T}$	$= \alpha_{b0} \cdot e^{-j\omega/\omega_T}$
$\alpha_e = \frac{i_c}{i_b}$	$\alpha_e = \frac{\alpha_b}{1-\alpha_b}$	$\frac{\alpha_{b0}}{1-\alpha_{b0}+j\omega/\omega_T}$	$\alpha_e = \frac{\alpha_b}{1-\alpha_b}$	$\frac{\alpha_{b0}}{1-\alpha_{b0}+j\omega/\omega_T}$	$\frac{\alpha_{b0}}{e^{+j\omega/\omega_T} - \alpha_{b0}}$
$Z_i = \frac{U_{be}}{i_b}$	$Z_i = r_e \cdot \alpha_e / \alpha_{b0}$	$\frac{r_e}{1-\alpha_{b0}+j\omega/\omega_T}$	$Z_i = r_e \cdot (\alpha_e + 1)$	$\frac{r_e \cdot (1+j\omega/\omega_T)}{1-\alpha_{b0}+j\omega/\omega_T}$	$\frac{r_e}{1-\alpha_{b0} \cdot e^{-j\omega/\omega_T}}$

Fig 3.16 Two essentially different modeling concepts for the BJT (bipolar junction transistor). Additional refinements, such as emitter inductance, base resistance and depletion layer capacitors, are omitted.

- The diffusion capacitance concept uses a frequency independent current source that senses the current flow through a resistor. An additional (diffusion) capacitor decreases the gain with increasing frequency.
- The current transport concept uses a frequency dependent controlled source that senses the current through a resistor. This transfer is usually a mix of delay and a low-pass transfer. Capacitors are not required for modeling the frequency response.

The current flow through r_e is proportional to the voltage between base and emitter. Therefore, many authors prefer the well-known hybrid- Π variant of the T -model that relies on diffusion capacitance.

The well-known hybrid- Π models, in which (diffusion) capacitance dominates the frequency response of bipolar transistors, are generally considered to be particularly suitable for describing theirs (linear) small-signal behavior [406]. Moreover, it can be concluded from microwave textbook discussions [301] that this assumption holds for multi-GHz frequencies. In our experience, BJT models dominantly based on (diffusion) capacitance become inadequate for the job for frequencies in excess of 10% of the transition frequency f_T .

When capacitive effects characterize the dominant frequency response, impedance z_e in figure 3.10 must be a real impedance that becomes capacitive for frequencies approaching the transition frequency f_T . Figure 3.8c illustrates that the virtual circuit parameter z_e is *inductive* rather than capacitive. From this we conclude that diffusion

capacitance is merely the result of first order approximations of BJT models up to $f_T/10$ then an adequate parameter for describing wideband transistor aspects.¹¹

In our experience, the *current transport factor* is the most convenient parameter to model the dominant frequency response of bipolar transistors in an adequate but simple way. This is the virtual parameter of α_b in figure 3.10, as has been extracted from measured data in section 3.2.2 and 3.2.3. The excellent fitting in figure 3.8 of α_b and α_e demonstrates the validity of this concept up to f_T .

Figure 3.15 illustrates the elementary modeling concepts for BJT's. The model in figure 3.15b is based on (diffusion) capacitance c_e and the model in figure 3.15c on a current transport function α_b . Additional refinements, such as base resistance, emitter resistance¹² and depletion layer capacitors are omitted to simplify this inter-comparison.

Diffusion capacitance concept

When the diffusion capacitance is a dominant effect, the virtual circuit impedance z_e must be shunted with a capacitor c_e in order to affect the frequency response significantly. Using the well-known relations $\omega_c \approx 1/(r_e \cdot c_e)$, and $r_e = (kT/q \cdot I_c)$, diffusion capacitance of $c_e \approx 5$ pF would have been expected based on transition frequency ($f_T = 3.5$ GHz) and bias current ($I_c = 3$ mA).

Figure 3.8c illustrates that z_e is *inductive* instead of capacitive. This 1.9 nH series inductor is probably caused by the emitter bond wire, and models a small parasitic effect. When some capacitive effect is noticed above f_T then its value is significantly smaller than the presumed diffusion capacitor c_e . In fact, we do not observe any dominant capacitance effect at all!

We verified this observation for various BJT's, using the s-parameters specified by the manufacturer¹³. The emitter impedance z_e is significantly inductive for all these devices, with an exception when the bias current I_c is less than 1 mA. When using relatively small bias currents, the value of r_e is relatively high. In these situations, a series inductor L_e is too small to be noticed below the transition frequency, and a small shunt capacitor dominates the frequency response. As a result, z_e is weakly capacitive for small bias currents due to the depletion layer capacitance.

From this observation, it is concluded that the junction capacitance between base and emitter is the most significant capacitor, which is usually very small and roughly independent of the bias current. As a result, the validity of BJT models based on diffusion capacitance concepts is restricted to frequencies below $f_T/10$.

¹¹ Naturally, improved fittings may result when fitting transistor models using more nodes and elements then is used in the equivalent circuit of figure 3.10. This has not been investigated because we prefer the simplest models.

¹² The actual value of r_e is somewhat higher then expected from predictions on the collector bias current. We observed an additional value ($r_{e0} \approx 1.3\Omega$) for the BJT examples in figure 3.8 and 3.12. This resistance remains roughly constant when the bias current is varied, as is demonstrated in figure 3.12. We conclude that this offset value is caused by emitter series resistance.

¹³ Philips: BF547, BF747, BFQ67, BFR92a, BFR93a, BFR106, BFR505, BFR520, BFR540, BFS17a, BFT25, BFT25a, BFT92, and BFT93.

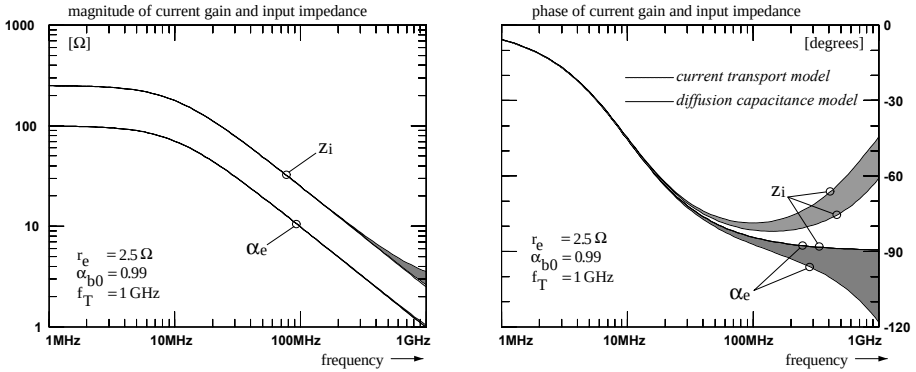


Fig 3.16 Simulated differences between BJT models dominantly based on the diffusion capacitance concept and on the current transport concept, as defined in figure 3.15. Either concepts predict similar magnitude responses however the phase prediction differs significantly for frequencies above $f_T/10$. The shaded area's demarcate the ranges for α_e and z_i when the current transfer in α_b varies from pure delay to first order low-pass transfer. The actual current gain is usually a mix of both transfer functions.

Current transport concept

The BJT models we proposed in section 3.2.2 is fundamentally different. The basic concept, as shown in figure 3.15c, assumes all elements to be frequency independent except for the current transport factor α_b . This concept is known from the very beginning of the transistor [205: page 190], but dropped into disuse.

For a wide range of BJT's, this concept is valid up to f_T . It is as simple as the commonly used diffusion capacitance concept, which is convenient for synthesis purposes.

Figure 3.16 shows simulated plots of current gain α_e and input impedance z_i . to compare either models. For frequencies up to f_T , the magnitude responses $|\alpha_e|$ are practically equal for both models. The same applies for $|z_i|$. Differences are mainly restricted to the phase response for frequencies above $f_T/10$.

The actual phase shift in α_e is higher then predicted from models based on diffusion capacitance. This difference is essential when simulating wideband feedback amplifiers since linear models based on diffusion capacitance concepts may incorrectly predict stable operation of oscillatory amplifiers.

3.3.3. Conclusions

We demonstrated that the well-known hybrid- Π model, in which (diffusion) capacitance dominates the frequency response of BJT's, becomes inadequate for frequencies in excess of 10% of the transition frequency f_T . This is accentuated by the fact that the dominant behavior of the virtual emitter impedance z_e is inductive instead of capacitive, for various wide band transistors biased between 1 mA and 20 mA. An improved linear BJT model has been developed, applicable up to f_T and suitable for synthesis. In this model, the current transport factor α_b dominates the frequency response. This improved model is essential when designing wideband feedback amplifiers, since it correctly predicts unstable operation of oscillatory amplifiers.

Chapter 4

Formal description of feedback

The design of feedback amplifiers involves the analysis and synthesis of the feedback loop to prevent oscillatory behavior. Various manual feedback analysis methods are developed since the very beginning [401,403], using loopgain and other well-known concepts. In addition, various automated circuit analysis methods are available since the development of circuit simulators.

Feedback analysis differs from circuit analysis. It is aimed at simplifying the transfer description of the signal flow in the feedback loop. This simplification is crucial for handling the complexity of feedback synthesis, as is discussed in chapter 5.

Restriction of conventional methods

Manual feedback analysis methods are adequate for relatively simple feedback networks at relatively low frequencies, e.g. below 1% of the transition frequency f_T of the individual transistors. The more sophisticated the feedback loop of a wideband amplifier is, the more its feedback analysis relies on automated circuit analysis.

Automated circuit analysis, using transistor models with estimated parameter values, may be adequate over a wider frequency interval. As a rule of the thumb, this applies when designing second order¹ feedback loops, up to 10% of f_T . The higher the feedback order and the wider the frequency band of interest, the more crucial a reliable circuit analysis will be.

The previous chapters 2 and 3 discussed in detail how to characterize the transfer of semiconductor devices, such as transistors. Furthermore, it was discussed how to extract adequate models from two-port measurements that take full account for relevant parasitic effects. These basic provisions facilitate an adequate circuit analysis, preferably with a simulator.

Wideband feedback design is limited by parasitic effects that are dominant. It requires a feedback analysis that takes full account for these effects. This holds especially when dealing with the stability of feedback loops, which is one of the main problems in wideband feedback design.

So far, we found no more than approximating methods in the literature to analyze feedback loops. Furthermore, there is in general no answer to the question of dealing with parasitic effects that modifies the dominant behavior of the feedback loop.

A novel approach

This chapter provides mathematical tools, to analyze feedback loops and their stability. It starts from a generalized feedback analysis that fully accounts for parasitic effects. Similarly to conventional synthesis methods, the proposed method is focussed on loop gain, however the transfer function is evaluated using multi-port methods.

¹ The term feedback *order* is adopted from filter theory, and refers to the dominant order of the loop-gain transfer-function. *Higher order* feedback is reserved for second order loops, third order loops or higher.

Furthermore, the proposed method extracts pole-zero representations for this loop gain, and converts these intermediate results into the simplest pole-zero representation that fulfills the accuracy requirements.

The highlights of this chapter, which are developed in this study, are:

- Development of a generalized feedback theory, that takes full account for parasitic effects and that is suitable for implementation in circuit simulators.
- Development of a new simple and robust algorithm that extracts the poles and zeros of a transfer function, for reliable pole-zero analysis. This was essential to the applicability of the feedback theory proposed here.
- Development of automated deflation algorithms that reduce the transfer order of the loop gain into its dominant form.
- Exact loop gain calculation, without ignoring any (linear) parasitic effects, by two specific cuts in the loop. The loop gain is derived from the two-port parameters of both the isolated parts of the loop.
- Accurate approximation of loop gain, by one cut in the loop at an arbitrary location. The loop gain is derived from the two-port parameters of the opened loop.
- Development of manual analysis methods for the contribution of zeros in the right half plane (RHP) to the overall loop transfer function.

4.1. Superposition analysis of feedback amplifiers

A linear circuit description using N-port matrix parameter is an exact representation. A tabular format, with numerical values for each frequency and node of interest is quite suitable for measurement and (automated) analysis. A basic drawback of this method is its large number of numerical data, which makes N-port matrix parameters inconvenient for manual transfer analysis and for manual compensation synthesis to realize stable feedback amplifiers.

The simplification of the problem is a successful approach in the design of feedback amplifiers. The elementary feedback model of Black's feedback patent [401], shown in figure 4.1, is based on this approach. Unfortunately, this model is too simple for wideband feedback design, because it requires transfer functions that are unilateral and not affecting each other,

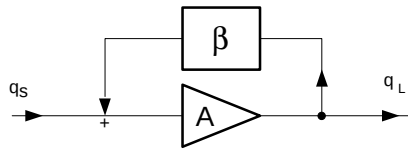


Fig 4.20 Flow diagram of Black's feedback model, which is too simple for wideband feedback design.

Nordholt [406] has described some extended models that are more suitable for this purpose. He formalized the superposition feedback analysis of Cherry and Hooper [405] with the description of the *superposition model*. This superposition model facilitates an exact representation of a feedback amplifier.

Further, the referred publication has reduced this superposition model to the *asymptotic gain model* which is simpler and therefore more convenient for manual analysis and synthesis. As a result, the synthesis is focussed on this asymptotic gain model and does not describe in detail how to extract all superposition parameters of an arbitrary circuit. This section generalizes the theory of the superposition model.

4.1.1. Superposition model parameters

In a well-designed feedback amplifier, the source signal is amplified by one or more cascaded amplifier stages and then fed to a load impedance. This enables a designer to designate a forward signal flow, and a feedback flow.

Figure 4.2 symbolizes this signal flow for an arbitrary feedback circuit. Source and load impedance are included in figure 4.2, because they affect the transfer and the stability of a feedback amplifier.

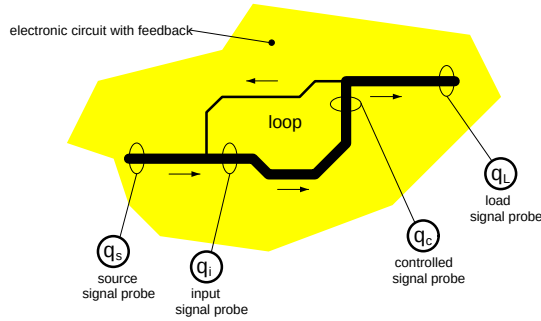


Fig 4.20 Signal flow in an arbitrary feedback amplifier, including the source and load. The four probes sense the signal flow through a reference plane, including voltage, current, wave, charge, or any other related quantity.

A superposition feedback analysis of the circuit in figure 4.2 requires that the signal flow is sensed by probes at four locations. The different probes may sense different quantities, including voltage, current, wave, charge or any other related quantity. It is the designer's choice what kind of signal is sensed, and as long as the signal dimension is undefined, they are referred as q_s , q_i , q_c and q_L .

The probes do not sense the signal at a single circuit node but through a reference plane. This means that a voltage-probe senses between a node and ground, a current-probe senses the current flow from one node to another, a wave-probe senses both voltage and current to reconstruct the wave, etcetera.

A superposition feedback analysis simplifies the circuit analysis as follows:

- It cuts the circuit in figure 4.2 at the four reference planes, associated with q_s , q_i , q_c and q_L , to split the circuit in (1) a one-port *source*, (2) a one-port *load*, (3) a two-port *forward amplifier* and (4) a four-port *feedback network* in which all remaining components resides.
- Then it reconstructs the original signal flow in the feedback network by external current injection in port q_s and q_c and proper termination of all the four ports with specific impedances. According to the superposition principle, the relation between the signal quantities has the following general matrix form:

$$\begin{bmatrix} q_L \\ q_i \end{bmatrix} = \begin{bmatrix} \rho & v \\ \xi & \beta \end{bmatrix} \cdot \begin{bmatrix} q_s \\ q_c \end{bmatrix}$$

- Next it reconstructs the original signal flow in the forward amplifier by external current injection in port q_i and proper termination of all the two ports with specific impedances. The relation between the resulting signal quantities has then the following general (matrix) form:

$$[q_c] = [A] \cdot [q_i]$$

Since the signal flow in the individual pieces is made identical to that of the overall circuit, the analysis of the individual pieces facilitates a faithful reproduction.

The superposition feedback analysis includes the presence of four signal quantities that are interrelated by five superposition parameters. We will refer to them by the following names² :

$q_c = A \cdot q_i$	$q_s = \text{source quantity}$	$A = \text{forward gain}$
$q_L = \rho \cdot q_s + v \cdot q_c$	$q_L = \text{load quantity}$	$\beta = \text{feedback factor}$
$q_i = \xi \cdot q_s + \beta \cdot q_c$	$q_i = \text{input quantity}$	$\xi = \text{input conversion}$
	$q_c = \text{controlled quantity}$	$v = \text{output conversion}$
		$\rho = \text{forward leakage}$

Fig 4.3 Superposition (model) parameters.

One extraction method of the five superposition parameters ρ , v , ξ , β and A starts from a full matrix description of the individual N-ports. The required terminations are then extracted from all these matrix parameters and the five superposition parameters are evaluated from these intermediate results. Since a full matrix description of an N-port requires a set with N^2 matrix parameters, this might be an inconvenient approach.

Another method starts from the observation that all source, load and termination impedances have specific values during the superposition analysis. This means that there is no need to evaluate all N-port matrix parameters. As a result, it will not isolate parts from the circuit in terms of individual N-ports but in terms of *ideal* sources, loads and amplifiers.

This approach yields the superposition *model*, and the block diagrams in figure 4.4 show different examples of this approach. The concept 'ideal' is closely related with the chosen dimension of the various signals. This means that output impedance of ideal sources is *zero* for a voltage source, *infinity* for a current source and *matched* for a wave source. Similar, the input impedance of voltage, current or wave probes are respectively infinity, zero or matched.

Figure 4.4a shows an example of the superposition model, in the special case that all signals of interest are voltages. As a result, it includes: an ideal voltage source with *zero output impedance* to generate q_s , an ideal voltage probe with *infinity input impedance* to sense q_L , and an ideal voltage to voltage amplifier with *infinity input impedance* to sense q_i and *zero output impedance* to generate q_c . All impedances that makes the actual source, load and amplifier non-ideal are embedded in a common feedback blackbox.

Figure 4.4b shows an example of the same feedback amplifier, in the case that the current is analyzed instead of the voltage. Using proper Thevenin to Norton transformations the physical source, load and amplifier are modified into ideal current sources and probes. As a result, it includes an ideal current source with *infinity output impedance* to generate q_s , an ideal current probe with *zero input impedance* to sense q_L , and an ideal current to current amplifier with *zero input impedance* to sense q_i and

² The names that are chosen in this thesis are reconciled with their physical interpretation. Alternative names that were found in the literature [406] for the transfer functions A , ε and v are *reference variable*, *input signal loss* and *output signal loss*.

infinity output impedance to generate q_c . All remaining impedances reside in a common feedback blackbox, which differs from that in figure 4.4a.

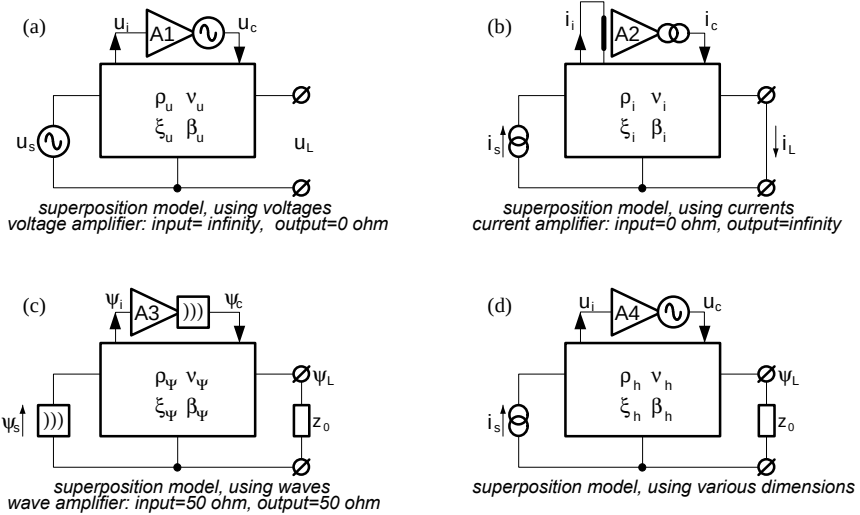


Fig 4.20 Examples of different implementations of the superposition model of feedback amplifiers, using signals with different dimensions. In these block diagrams all sources, probes and amplifiers are ideal by definition. This means, for instance, that A2 is an ideal current amplifier with zero input impedance and infinity output impedance. The four-port represents the feedback blackbox, in which all remaining components of the circuit reside.

Figure 4.4c shows an example of the same feedback amplifier, in the case that all signals of interest are waves. Similar to the example in figure 4.4b, the source load and amplifier are transformed to ideal components that are perfectly matched to a chosen Z_0 (e.g. $50\ \Omega$).

In the example of figure 4.4d, all signals of interest have different dimensions.

It is irrelevant whether the forward amplifier is a single amplifier stage or a cascade of stages. The superposition feedback analysis models this cascade by an ideal forward amplifier surrounded by input, output and shunted impedances. It then extracts these impedances from the forward amplifier and embeds them in the feedback network. All remaining circuit components reside in a common multi-terminal network, the *feedback network*.

4.1.2. Superposition flow parameters

The superposition model originates from well-considered mathematical manipulations with the feedback circuit, however that approach estranges the model of a physical interpretation. The superposition flow diagrams in figure 4.5 represent the equations of the superposition model in a graphic way to relax the interpretation of the superposition parameters.

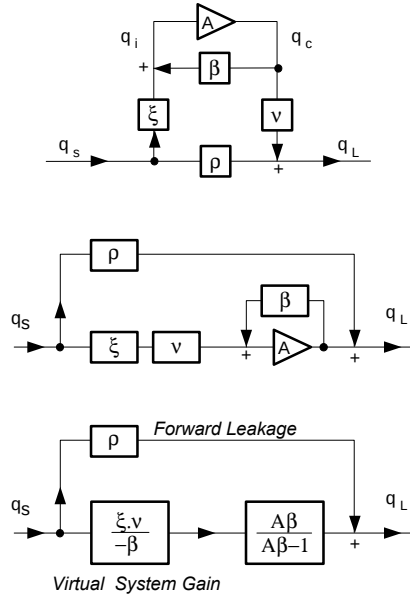


Fig 4.20 Three identical flow diagrams of the superposition model. Some transfer functions are grouped to emphasize their influence on the overall result.

We define some additional transfer functions, and their associated names, because their physical interpretation is meaningful. The last flow diagram in figure 4.5 illustrates the meaning of these superposition flow parameters.

- The *loop gain* ($A\beta$) is a measure for the transfer of all amplifier stages and feedback attenuation in the feedback loop.
- The *virtual system gain* ($-\xi \cdot v / \beta$) is a measure for the system gain in the case of zero leakage ($\rho \rightarrow 0$) and infinity loop gain ($A\beta \rightarrow \infty$). When the loop gain reduces below the unity gain value, then the virtual system gain differs significantly from the realized system gain.
- The *asymptotic system gain* ($\rho - \xi \cdot v / \beta$) is a measure comparative to the virtual system gain, without neglecting the forward leakage.
- The *realized system gain* is the transfer that is actually achieved.

The *forward leakage* (ρ) is the superposition model parameter that was defined in section 4.1.1. It is a measure for the isolation between input and output. When the forward gain reduces to zero ($A \rightarrow 0$), then the gain of the overall system becomes equal to this leakage value.

<i>loop gain:</i>	$\dot{H}$	$= A \cdot \beta$
<i>forward leakage</i>	A_{t0}	$= \rho$
<i>virtual system gain</i>	$A_{v\infty}$	$= -(\xi \cdot v / \beta)$
<i>asymptotic system gain</i>	$A_{t\infty}$	$= \rho - (\xi \cdot v / \beta)$
 <i>realized system gain:</i>		
$A_t = (A_{t\infty} - A_{t0}) \cdot \left(\frac{A\beta}{A\beta - 1} \right) + A_{t0} \approx A_{t\infty} \cdot \left(\frac{A\beta}{A\beta - 1} \right)$		

Fig 4.6 Extraction of superposition flow-parameters from y-parameters.

The superposition model is an exact representation of the feedback amplifier, and does not neglect parasitic effects. It is not required to embed *all* active devices in the forward gain block. Further, the choice of the dimension of the input quantity and the choice of the controlled quantity is arbitrary.

In the special case of a multi loop feedback configuration, where the controlled amplifier has more than two output nodes, it is sufficient to embed a *few* active devices in the forward gain block and to embed all remaining active and passive devices in the multi terminal feedback blackbox. As a result, the choice of the superposition parameters is ambiguous.

This will be illustrated with an example. Consider a situation in which the input and output nodes of the forward gain block are chosen. Then a variation in the dimension of the chosen input and controlled quantity will not affect the forward leakage, the loop gain and the virtual system gain. To prove this, change the input quantity q_i to $q'_i = \alpha_i \cdot q_i$ and change the controlled quantity q_c to $q'_c = \alpha_c \cdot q_c$. Then the modified superposition parameters are related as follow:

$$\begin{array}{lll}
 q_c = q'_c / \alpha_c & A = A' \cdot \alpha_c / \alpha_i & (A \cdot \beta) = \text{invariant} \\
 q_i = q'_i / \alpha_i & \beta = \beta' \cdot \alpha_i / \alpha_c & \\
 & \xi = \xi' \cdot \alpha_i & (\xi \cdot v / \beta) = \text{invariant} \\
 & v = v' / \alpha_c & \\
 & \rho = \rho' & \rho = \text{invariant}
 \end{array}$$

Although the choice of the nodes and the dimension of the internal signal is arbitrary from a mathematical point of view, the choice is most handy when all active devices are embedded in the forward amplifier block. This choice minimizes the gap between conventional feedback analysis and superposition feedback analysis.

4.1.3. Example of the extraction of superposition parameters

A simple example how to evaluate the superposition model is draw in figure 4.7. An electronic circuit with transimpedance feedback and a field effect transistor amplifies a current and generates a voltage for a load impedance. The associated superposition model is hybrid because source and load dimension are different.

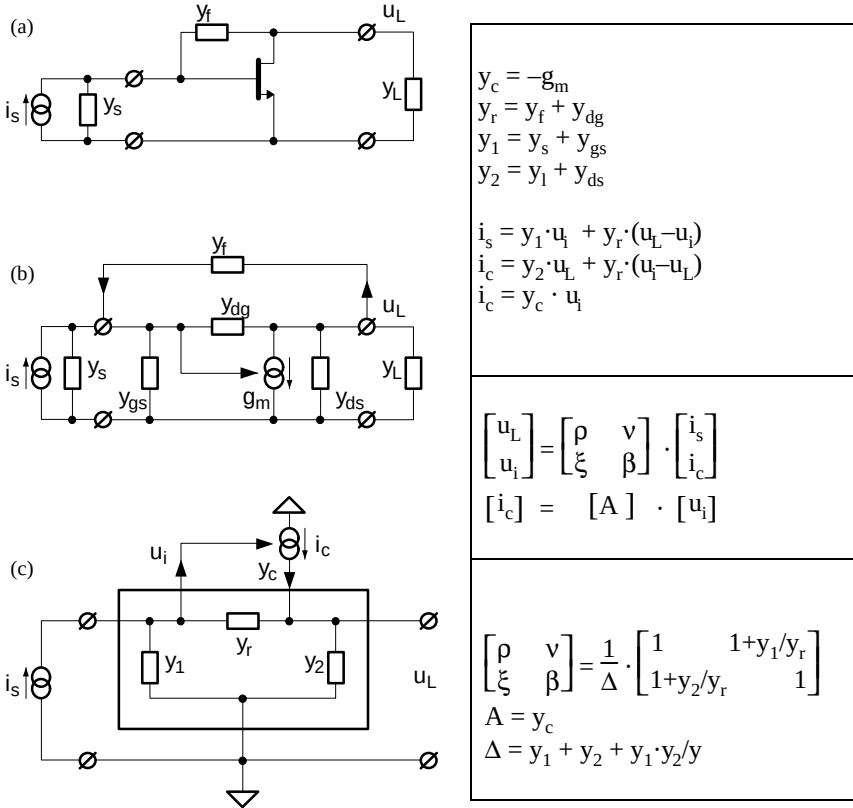


Fig 4.20 Example of the evaluation of the superposition model for transimpedance feedback. The electronic circuit (A) is modeled by a linear P-network (B). Then, all shunt impedances are combined (C) and some ideal (controlled) sources are isolated from the remaining elements.

The first step in the evaluation of the superposition model is the replacement of the active device by an appropriate circuit model. The example in figure 4.7b uses a Π -FET model with three impedances and a controlled source to represent the circuit of figure 4.7a.

The next step is to isolate the (controlled) sources from the rest of the circuit and to combine all remaining components in a common blackbox. The example in figure 4.7c combines all shunted impedances to reduce the complexity of the blackbox. The result is a distinct separation between feedback network, forward amplifier and source and load

quantities. Further, the figure summarizes the relation between circuit parameters and superposition parameters.

With some mathematical effort figure 4.7 results in:

$A = y_c$	$\frac{A\beta}{A\beta-1} = \frac{y_c}{y_c - y_1 - y_2 - y_1 \cdot y_2 / y_r}$
$\beta = \frac{1}{y_1 + y_2 + y_1 \cdot y_2 / y_r}$	$A_{t\infty} = \frac{-1}{y_r}$
$A\beta = \frac{y_c}{y_1 + y_2 + y_1 \cdot y_2 / y_r}$	$A_{t0} = \frac{1}{y_1 + y_2 + y_1 \cdot y_2 / y_r}$
$\xi \cdot v / \beta = \frac{(y_r + y_1) \cdot (y_r + y_2)}{y_1 + y_2 + y_1 \cdot y_2 / y_r}$	$A_t = \frac{1}{(y_1 - y_c) \cdot (y_2 + y_r) / (y_c + y_r) + y_2}$
$\rho = \frac{1}{y_1 + y_2 + y_1 \cdot y_2 / y_r}$	

Fig 4.8 Extraction of superposition flow parameters from y -parameters.

Note that $A_{t\infty}$ is in this example not equivalent to the desired system gain $(-1/y_r)$. To simplify this example, the parasitic feedback impedance y_{dg} (capacitance) is incorporated in the external feedback impedance (y_r). Section 4.2 describes a general applicable method.

4.1.4. Conclusions

In conclusion, the superposition model for an arbitrary feedback amplifier is described in general terms. The dimension of the signal flow of interest makes the superposition model more specific. This has been illustrated for signals such as voltages, currents and waves.

The superposition model originates from a pure mathematical definition, with five *superposition model parameters*. The interpretation of this model is relaxed using flow diagrams, and some transfer functions are grouped to reduce the number of parameters.

The superposition model is an exact representation of the feedback amplifier, however, the superposition parameters are not unique. The choice of the internal signal quantities and their dimensions is decisive. It has been demonstrated that a transformation in internal signal dimension will transform most superposition parameters. Some combined quantities remain invariant under this transformation. This holds for the loop gain ($A\beta$), the virtual system gain $(-\epsilon \cdot v / \beta)$ and the forward leakage (ρ).

The evaluation of all superposition parameters of a simple transimpedance amplifier has been demonstrated as example.

4.2. Superposition parameter calculation

The previous section described the superposition model in general terms. Feedback analysis, using the superposition model, requires the evaluation of all relevant superposition parameters for arbitrary feedback amplifier configurations.

Nordholt [406] has solved this by simplifying the analysis to the asymptotic gain model. Since this asymptotic gain model ignores the forward leakage (p), it is inadequate for wideband feedback design. The previous section described a restricted solution for a specific configuration.

This study has resulted in a generalized method to extract all superposition parameters, using N-port techniques. The feedback circuit is split in two sub circuits. The associated N-port parameters are evaluated with a circuit simulator and represented in a y-parameter format. The various superposition parameters are calculated using these y-parameters. This section describes the proposed generalized method.

4.2.1. Calculation of loop gain, forward gain and feedback factor

An arbitrary feedback circuit may contain one or more active devices, and thus may contain several non-ideal controlled sources with (parasitic) local feedback. From a mathematical point of view it is correct to choose one of these controlled sources as forward amplifier. The other controlled sources can be placed among the components in the feedback network. This emphasizes that the feedback network is not confined to passive components only; active components are also allowed.

From a physical point of view, an arbitrary choice is a very inconvenient approach. A clumsy choice for A and β will be compensated by inconvenient conversion values for ϵ and v . However, this approach estranges the model from its traditional physical meaning. The best choice, from a physical point of view, causes unity conversion at input and output ($\epsilon=1$, $v=1$, both with the associated dimension) and an equal dimension for forward gain (A) and system gain (A_L).

An example of an inconvenient choice for forward gain is the use of the controlled source in a *single* amplifier stage of a multi stage loop.

One of the highlights of this study is the idea that an arbitrary two-port is fully characterized by a virtual circuit with four elements. This was discussed in detail in section 2.3.2 when introducing virtual circuit parameters. Figure 4.9 shows an elaboration of this approach, that is more suitable for physical interpretation than methods that isolate the controlled source in a *single* amplifier stage.

The cascade of amplifier stages is represented by a two-port network, and the associated two-port y-parameters facilitate the extraction of the four elements of an equivalent circuit. As a result, the extracted source represents a cascade of various active devices with a *single* controlled source. This approach facilitates the evaluation of the superposition parameters in a way, similar to the example of subsection 4.1.3.

The method proposed here can be implemented simply in a circuit simulator that is conversant with N-ports as a compound circuit element.

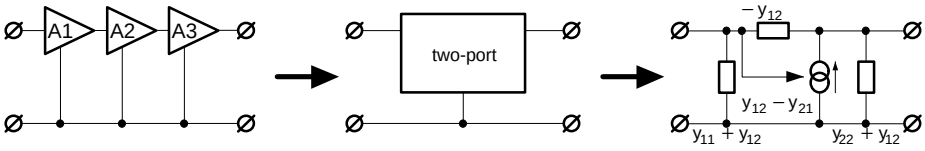


Fig 4.20 The ideal controlled source, embedded in a cascade of amplifier stages, is de-embedded by a transformation of two-port Y-parameters into P-model-parameters. Note that the circuit representation of the P-model-parameters holds for any arbitrary two-port and is not restricted to FET's.

Figure 4.10 shows a more general approach for the loop gain calculation with a circuit simulator. Since loop gain $A\beta$ is dimensionless and independent on the chosen dimension of A or β , the dimension of transconductance is used as a temporary choice.

- The simulation of an arbitrary circuit starts with defining a path of forward signal flow and isolates four sub-circuits, as shown in figure 4.10a. These sub-circuits are (1) source impedance, (2) load impedance, (3) a four-port feedback network and (4) a two-port forward amplifier. The multi-port parameters of each sub-circuit are subsequently evaluated with a circuit simulator, for instance as described in section 2.2.4.
- Then, the multi-port parameters of sub-circuit 1, 2 and 3 are combined and reduced for extracting additional two-port parameters with node n_4 as input port node and node n_3 as output port node. As is shown in figure 4.10b, the circuit has been split into two networks: one forward gain two-port and a reduced feedback two-port. The required algorithms are discussed in section 2.2.3 and 2.2.4.
- Next, these two-port parameters are transformed in (Π model) virtual circuit parameters for generating the virtual circuits as shown in figure 4.10c. The required algorithms are discussed in section 2.3.2.
- Subsequently, the simulation 'moves' the equivalent passive circuit elements from the forward gain two-port to the feedback two-port for de-embedding the controlled source parameter y_c . Figure 4.10d shows how the equivalent loop model is simplified using this approach for representing the feedback loop in a standard format.
- The standard format of the feedback network facilitates the calculation of the superposition parameters and the loop gain in a simple way, similar to the example in section 4.1.3. The loop gain is calculated similar to the methods discussed for figure 4.7.

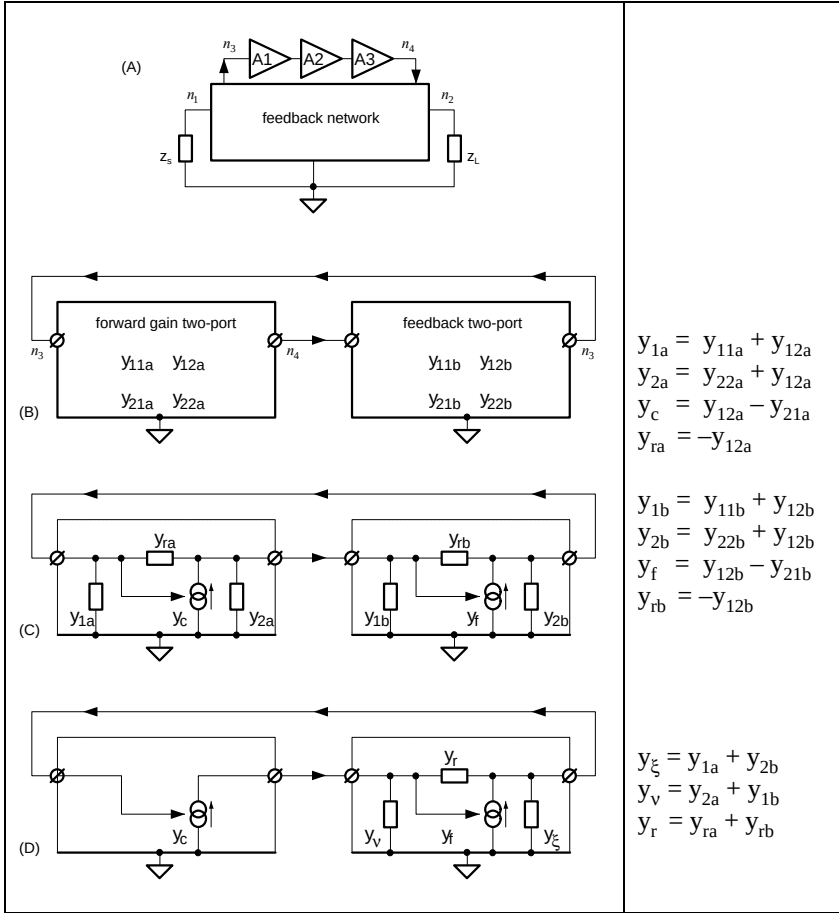


Fig 4.20 Intermediate calculation steps for the evaluation of the loop gain A_b . The source and load impedances are enclosed in the feedback blackbox (A), and both the amplifier blackbox and the feedback blackbox are represented by two-port networks (B). The y-parameters of these two-ports are transformed into virtual circuit-parameters (C), (see section 2.3.2). The equivalent input, output and shunt impedances are stripped from the amplifier and combined with the feedback network (D). The loop gain is finally evaluated from this simplified model.

- Next, the dimension of the forward gain is chosen, which is most meaningful when its dimension equals that of the realized system gain. The default choice in figure 4.10 is $[1/\Omega]$ because of the voltage controlled current source y_c . The transformation of the forward amplifier into another controlled source requires additional impedance at input and/or output. The input admittance y_{1a} and the output admittance y_{2a} are appropriate default choices. The forward gain A is one of the following list:

$$\begin{aligned}
 \text{voltage to current:} & \quad A_y = y_c \\
 \text{current to current:} & \quad A_i = y_c/y_{1a} \\
 \text{voltage to voltage:} & \quad A_u = y_c/y_{2a} \\
 \text{current to voltage:} & \quad A_z = y_c/(y_{1a} \cdot y_{2a})
 \end{aligned}$$

- Finally, the feedback factor β is determined by the quotient of loop gain and forward gain.

The above procedure is suitable for manual calculation however it is in its current form, with equivalent circuit elements, too cumbersome to implement in a circuit simulator. More appropriate is a direct evaluation of the required quantities from the two-port parameters, for instance from the y-parameters. With some mathematical effort the following expressions can be obtained for loop gain and some basic forms of the forward gain:

forward gain (I/U)	$A_y = y_c$	$y_\xi = y_{11a} + y_{22b} + y_{12a} + y_{12b}$
forward gain (I/I)	$A_i = y_c/y_{1a}$	$y_v = y_{11b} + y_{22a} + y_{12a} + y_{12b}$
forward gain (U/U)	$A_u = y_c/y_{2a}$	$y_r = -(y_{12a} + y_{12b})$
forward gain (U/I)	$A_z = y_c/(y_{1a} \cdot y_{2a})$	$y_c = y_{12a} - y_{21a}$
		$y_f = y_{12b} - y_{21b}$
		$y_{1a} = y_{11a} + y_{12a}$
		$y_{2a} = y_{22a} + y_{12a}$
loop gain	$A\beta = \frac{y_c \cdot (y_f + y_r)}{y_\xi \cdot y_v + y_r \cdot (y_\xi + y_v - y_f)}$	

Fig 4.11 Extraction of loop gain and forward gain from y-parameters.

The previous approach is valid for grounded amplifier stages only. In the case that the input or output amplifier stage has floating external nodes, then the N-port description of the forward amplifier requires at least three ports. These configurations are not considered here, however the loop gain has to be calculated in a similar way.

4.2.2. Calculation of asymptotic and virtual gain and forward leakage

The asymptotic system gain is the system gain under the hypothetical situation that the loop gain goes to infinity ($\hat{H} = A\beta \rightarrow \infty$). The same applies for the forward leakage, under the situation that the loop gain goes to zero ($\hat{H} = A\beta \rightarrow 0$). The value of realized system gain is somewhere between these extreme values. The relation between loop gain variation γ and associated variable system gain $A_t(\gamma)$ is as follows:

$$\text{variable system gain:} \quad A_t(\gamma) = (A_{t\infty} - A_{t0}) \cdot \left(\frac{\delta \cdot \hat{H}}{\delta \cdot \hat{H} - 1} \right) + A_{t0}$$

Since the loop gain $\hat{H}$ is known from previous calculations, the evaluation of these transfer functions is feasible by system gain calculations under variation of loop gain.

forward leakage:	$A_{t0} = A_t(0)$
asymptotic system gain:	$A_{t\infty} = (A_t(1) - A_t(0)) \cdot \left(\frac{\frac{0}{H}-1}{\frac{0}{H}} \right) + A_t(0)$
virtual system gain:	$A_{\infty} = A_t(1) - A_t(0)$

Fig 4.12 Extraction of forward leakage, asymptotic system gain and virtual system gain from y -parameters.

Arithmetic variation of loop gain requires the isolation of the forward amplifier from the circuit as a two-port. Variation of the transconductance y_c of the equivalent circuit representation of this two-port is similar to the variation of loop gain.

A successful mathematical approach for circuit simulators is the evaluation of the y -parameters of the two-port, the modification of the y_{21} parameter, and then the evaluation of the overall system transfer with this modified two-port. Linear variation of loop gain requires a variation of $y'_{21} = \gamma y_{21} - (1-\gamma) y_{12}$.

Figure 4.13 shows the circuit equivalent of this procedure. Forward leakage calculation requires a modification of $y'_{21} = y_{12}$ to perform zero loop gain.

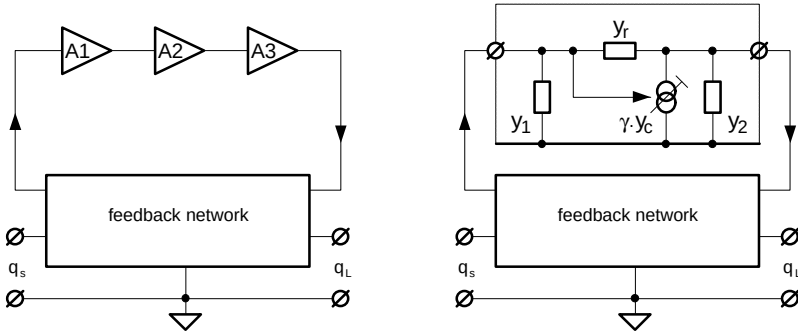


Fig 4.20 Arithmetic variation of loop gain by modification of the y_{21} parameter of the cascade of amplifier stages. When the controlled source transfer y_c has to change to $y'_c = \gamma y_c$, then the y_{21} parameter has to change to $y'_{21} = \gamma y_{21} - (1-\gamma) y_{12}$

4.2.3. Conclusions

In conclusion, a generalized method is demonstrated for extracting all superposition parameters from arbitrary feedback circuits. The application of the proposed method is primarily intended for usage with circuit simulators, equipped with N-port facilities.

The designer splits the circuit in two sub-circuits and defines the dimension of the internal and external signals. The simulator provides the five superposition model parameters.

4.3. Loop gain estimation using single-cut methods

An *exact* superposition analysis requires an exact calculation of the superposition parameters. The wider the passband of the amplifier, the more the analysis must rely on exact values. When future circuit simulators give full support to superposition analysis, then exact analysis is always preferred.

Sometimes, an *estimated* superposition analysis is useful, for instance when the relevant parameters are directly obtained from two-port measurements. It requires that internal nodes of the circuit under test are made available for external measurement, such as a modified circuit with opened feedback loop. Further, estimated analysis might be useful to obtain global insight on stability performance with minimal calculation effort. This holds especially when appropriated (software) tools are not available.

Estimated feedback analysis is primarily focused on loop gain calculation since most stability problems are mainly related on loop gain. It may serve as an intermediate step between conventional analysis and exact superposition analysis.

This section describes mathematical methods to estimate the loop gain. This estimated result will be referred in this text by the name: *open* loop gain.

Further, this section links the exact definition of loop gain with the estimated methods that are proposed here and with conventional loop gain estimations. This facilitates the assessment of errors originating from open loop gain and conventional methods.

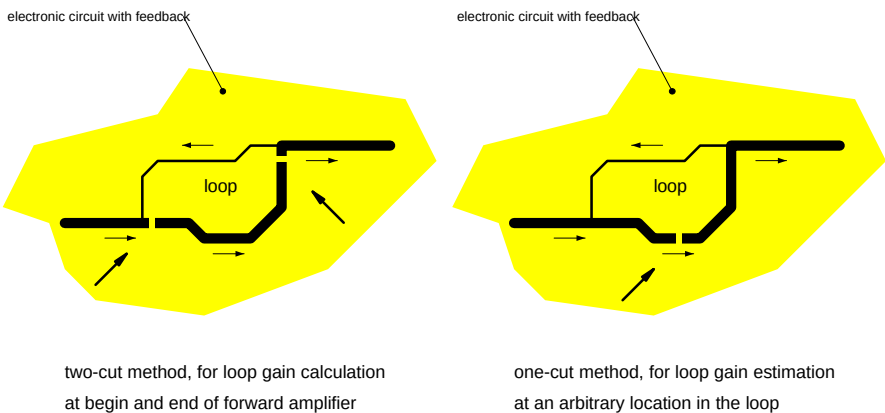


Fig 4.20 Signal flow in a feedback amplifier. Two cuts in the loop will isolate the forward amplifier from feedback network and facilitate an exact calculation of the loop gain. With a single cut in the loop, the determination of the loop gain is restricted to an approximation.

Figure 4.14 illustrates the basic idea, on the basis of signal flow in an arbitrary feedback amplifier.

With two cuts at the begin and end of the forward amplifier, this amplifier is isolated from the feedback network, which facilitates an exact loop gain calculation. Two cuts in the loop yields two individual two-ports and eight two-port parameters. An exact loop gain calculation, make use of all these parameters.

A single cut yields an single (open) loop two-port and four two-port parameters. This restricts the calculated result to a fair estimation of loop gain. It will be demonstrated that this estimation is independent of the position of the cut.

Although two-cut methods are preferred, due to their accuracy, one-cut methods are sometimes attractive because of their simplicity.

4.3.1. One-cut loop gain estimation using two-port parameters

Single cut methods start from the assumption that the transfer of the forward amplifier is close to unilateral at all frequencies of interest. This means that the forward transfer (gain) of that two-port is significantly higher than the reverse transfer (isolation). Note that this condition is not required for the *individual* amplifier stages; the more stages are used in the overall cascade, the better the overall transfer is unilateral.

The concept of *sufficient* isolation of the forward amplifier is related to the transfer of the feedback network. The feedback must dominate the isolation. In general, the more stages are used in the overall cascade, the lower the feedback is and the higher the isolation demands will be.

The higher the frequency of interest, the lower the isolation will be. As a result, the application of one-cut loop gain methods are mainly restricted to frequencies that are significantly lower than transition frequencies f_T of the transistors.

Figure 4.15 shows various representations of the opened loop.

Figure 4.15a shows the result from the two-cut method. Forward amplifier two-port and the feedback network are isolated from the circuit. Because the source and load impedance are embedded in the feedback network, it is represented here as two-port. Both two-ports represent the loop.

Figure 4.15b shows the two-port parameters using a virtual circuit representation. The element values are extracted from the Π -model virtual circuit-parameters, as was previously described in section 2.3.2. When all eight numbers from both two-ports are available, then an exact loop gain calculation is feasible.

Figure 4.15c shows the cascade of both two-ports using a virtual circuit representation. This two-port is the open loop two-port and the element values are derived from the two-port parameters from figure 4.15b.

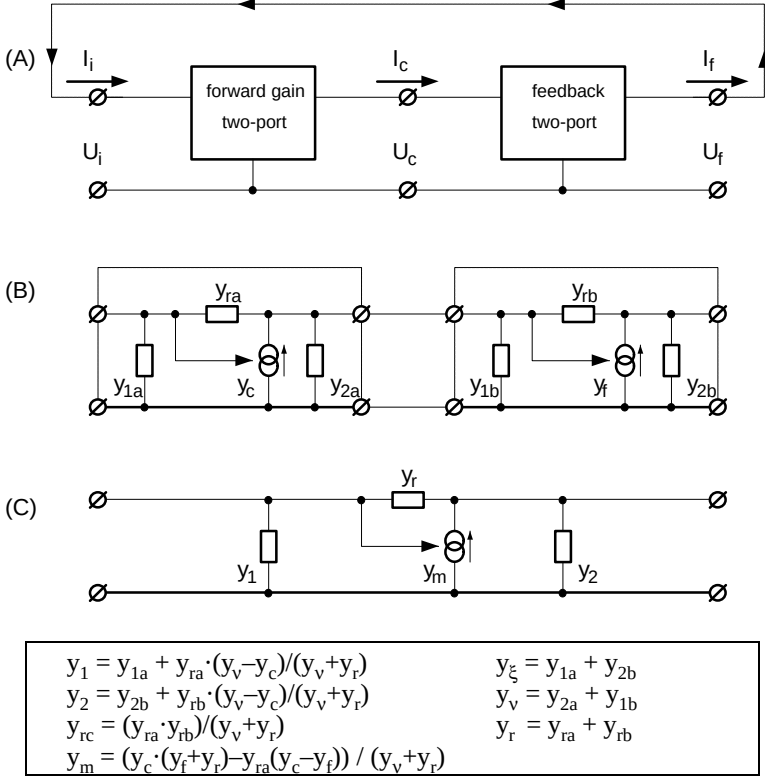


Fig 4.20 Approximation of the loop gain $\hat{H} = \mathbf{A}\mathbf{b}$ by the open loop gain. If the isolation of the forward amplifier is high ($y_{ra} \rightarrow 0$) then the open loop gain is close to the loop gain.

definition of open loop gain

The evaluation of the loop gain $\hat{H}$, from the individual two-port parameters of forward amplifier and feedback network, was derived section 4.2.1. We have found a fair estimation of $\hat{H}$, which is defined as $\hat{H} \stackrel{\text{def}}{=} (y_m) / (y_1 + y_2)$, that is evaluated from the open loop two-port parameters. It can be derived from the equations in figure 4.15 that both quantities are related as follows:

$$\begin{aligned} \text{loop gain:} \quad \hat{H} &= \frac{y_c \cdot (y_f + y_r)}{y_{\xi} \cdot y_v + y_r \cdot (y_{\xi} + y_v - y_f)} \\ \text{open loop gain:} \quad \hat{\hat{H}} &\stackrel{\text{def}}{=} \frac{y_c \cdot (y_f + y_r) - \delta}{y_{\xi} \cdot y_v + y_r \cdot (y_{\xi} + y_v - y_f) - \delta} = \frac{y_m}{y_1 + y_2} \\ &\quad \text{with } \delta = y_{ra} \cdot (y_c - y_f) \end{aligned}$$

When y_{ra} is small enough to ignore δ then the open loop gain $\hat{\hat{H}}$ is close to the loop gain $\hat{H}$.

relation with other two-port parameters

The definition of the open loop gain is based on the virtual circuit parameters (y_m , y_1 , y_2). However these two-port parameters are not always direct available. In practice, it is often handy to use s-parameters for measurement and y-parameters for circuit simulation, however any two-port parameter set applies. Therefore it is convenient to evaluate the open loop gain directly from arbitrary two-port parameters. Using the equations of figure 4.15, and the two-port parameter definitions of section 2.3.1, it can be demonstrated that the following relations hold for $\hat{H}$:

open loop gain for various two-port parameters:

$\hat{H} = \frac{y_m}{y_1 + y_2}$ $\hat{H} = \frac{y_{12} - y_{21}}{y_{11} + y_{22} + 2 \cdot y_{12}}$ $\hat{H} = \frac{z_{21} - z_{12}}{z_{11} + z_{22} - 2 \cdot z_{12}}$	$\hat{H} = \frac{h_{21} + h_{12}}{2 \cdot h_{12} - 1 - \Delta_h}$ $\hat{H} = \frac{1 - \Delta_a}{a_{11} + a_{22} - 2 \cdot \Delta_a}$ $\hat{H} = \frac{1 - \Delta_{abcd}}{A + D - 2 \cdot \Delta_{abcd}}$	$\hat{H} = \frac{s_{12} - s_{21}}{2 \cdot s_{12} + \Delta_s - 1}$ $\hat{H} = \frac{1 - \Delta_t}{t_{11} + t_{22} - 2 \cdot \Delta_t}$

$\Delta_a = a_{11} \cdot a_{22} - a_{12} \cdot a_{21}$
 $\Delta_h = h_{11} \cdot h_{22} - h_{12} \cdot h_{21}$
 $\Delta_{abcd} = A \cdot D - B \cdot C$
 $\Delta_s = s_{11} \cdot s_{22} - s_{12} \cdot s_{21}$
 $\Delta_t = t_{11} \cdot t_{22} - t_{12} \cdot t_{21}$

relation with current gain and voltage gain

The open loop gain $\hat{H}$ is a transfer function somewhere in-between current gain H_i and voltage gain H_u . Figure 4.16 demonstrates this by means of three circuits with identical open loop gain. Figure 4.16a is the virtual circuit representation of an open loop two-port. Figure 4.16b and 4.16c are different two-ports, but have in common that the open loop gain is equal for all three two-ports. The major difference is that the input impedance is moved to the output (or reverse), and that the shunt impedance is removed. The current gain for the two-port in figure 4.16b equals the open loop gain, while the same applies for the voltage gain in figure 4.16c. As a result, when $y_1 \approx 0$ then $\hat{H}$ approaches the voltage gain and when $y_2 \approx 0$ then $\hat{H}$ approaches the current gain.

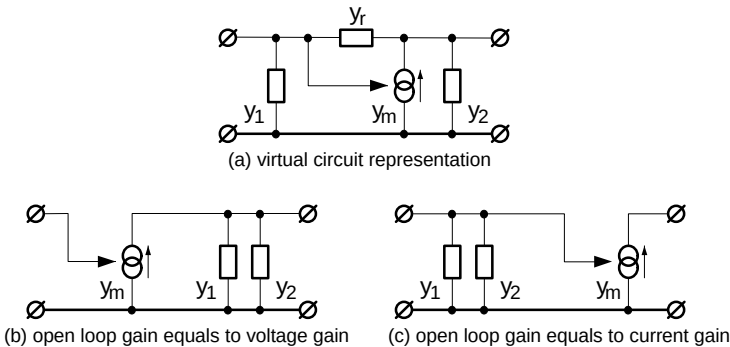


Fig 4.16 Three circuits with identical open loop gain.

position invariance

The location of the loop cut is irrelevant, as long as the cyclic order remains unchanged. This means that the open loop gain of a two-port cascade with order (1,2,3,4) equals that of the cascade with order (2,3,4,1) or (3,4,1,2) or (4,1,2,3). As a result, the value of the open loop gain is unique for a cascaded loop of arbitrary two-ports, and invariant of the cut location.

The property that open loop gain is *position invariant* is simply verified with a matrix representation of the individual two-ports. The simplest verification uses t-parameters, because the T-matrix of a cascade of two-port is equal to matrix product of all individual T-matrices. When the matrices T_a , T_b and T_c are representations of arbitrary two-ports, then the open loop gain of $T_a \cdot T_b \cdot T_c$ equals both the open loop gains of $T_b \cdot T_c \cdot T_a$ and $T_c \cdot T_a \cdot T_b$. This property holds for arbitrary t-parameter sets and does not require a unilateral behavior of the individual two-ports.

Figure 4.17 illustrates the benefits of this position invariance. The loop is cut at an arbitrary node in the signal path through the amplifier stages. The impedance value of input and output impedance of the separated amplifier stages is irrelevant.

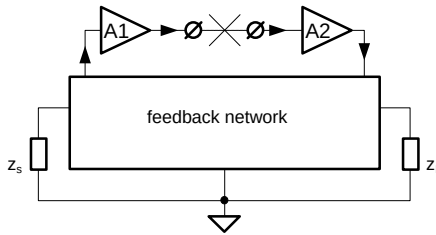


Fig 4.20 Approximation of the loop gain $\hat{H} = \hat{A}b$ by the measurement of the open loop gain $\hat{H}$. The loop is opened at an arbitrary location, and $\hat{H}$ is reconstructed from the measured two-port parameters. Note that $\hat{H}$ is a unique parameter of a loop of cascaded two-ports, as is described in the text.

4.3.2. One-cut loop gain estimation using open voltage and current gain

One-cut loop gain estimations, using two-port parameters, can be implemented simply in circuit simulators with two-port parameter facilities. This results in the open loop gain, that is an approximation for the exact loop gain. Manual loop gain calculation methods prefer the use of currents and voltages. Commonly used manual methods evaluate the open loop *voltage* gain or the open loop *current* gain, which are approximations of the open loop gain.

This subsection relates the open loop gain with open loop *voltage* gain and the open loop *current* gain, to assess whether these approximations are applicable or not.

Tuinenga [407: chapter 6.9] described an elegant way to define the open loop gain for voltages and for currents. The basic principle is outlined in figure 4.18. An external source injects a signal in the loop and a voltage or current ratio meter senses the injected

signals. The current ratio $H_i = i_y/i_x$ is referred in this text as *open loop current-gain* and the voltage ratio $H_u = u_y/u_x$ as *open loop voltage-gain*.

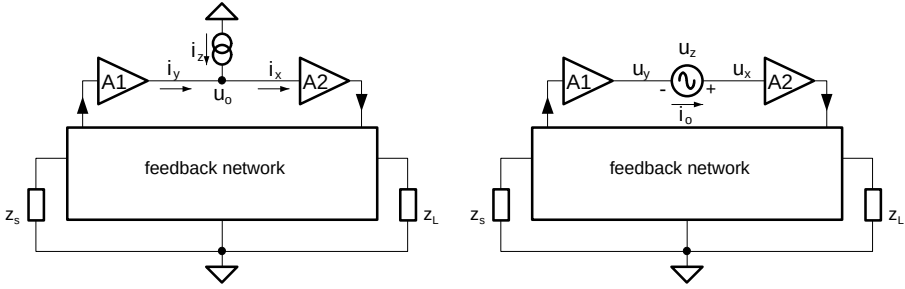


Fig 4.20 Approximation of the open loop gain $\hat{H}$ by measurement of the open loop current-gain $H_i = i_y/i_x$ or the open loop voltage-gain $H_u = u_y/u_x$

The calculation of the injected signal ratio of the open loop two-port is facilitated by a y-parameter representation. Since these parameters (y_{11} , y_{21} , y_{12} , y_{22}) are related with the virtual circuit parameters (y_m , y_1 , y_2 , y_r) of figure 4.15t can be demonstrated that the following conditions hold:

$$\text{open loop current gain} \quad \begin{bmatrix} i_x \\ -i_y \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \cdot \begin{bmatrix} u_0 \\ u_0 \end{bmatrix} \Rightarrow H_i \stackrel{\text{def}}{=} \frac{i_y}{i_x} = -\frac{y_{22} + y_{21}}{y_{11} + y_{12}} = \frac{y_m - y_2}{y_1}$$

$$\text{open loop voltage gain} \quad \begin{bmatrix} i_0 \\ -i_0 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \cdot \begin{bmatrix} u_x \\ u_y \end{bmatrix} \Rightarrow H_u \stackrel{\text{def}}{=} \frac{u_y}{u_x} = -\frac{y_{11} + y_{21}}{y_{22} + y_{12}} = \frac{y_m - y_1}{y_2}$$

The value of the reverse isolation $y_r = -y_{12}$ is irrelevant in this measurement setup, because the associated feedback impedance is virtually shorted during the measurements. The current or voltage ratio approaches the open loop gain $\hat{H}$ when the magnitudes of the ratio (y_1/y_2) differ significantly from one. It can be demonstrated that H_i and H_u are as follows related to $\hat{H}$:

$$\begin{aligned} H_i &= \hat{H} + (y_2/y_1) \cdot (\hat{H} - 1) & \text{if } |y_2/y_1| \ll 1 & \text{ then } H_i \approx \hat{H} \\ H_u &= \hat{H} + (y_1/y_2) \cdot (\hat{H} - 1) & \text{if } |y_2/y_1| \gg 1 & \text{ then } H_u \approx \hat{H} \end{aligned}$$

Both ratio measurements are poor approximations of the open loop gain, however combining H_i and H_u facilitates the evaluation of the correct value. It can be demonstrated that the following relation holds:

$$\hat{H} = \frac{H_i \cdot H_u - 1}{H_i + H_u - 2}$$

The open loop current gain or open loop voltage gain dominates in the above expression when the magnitudes of input and output impedances are significantly unbalanced. The *lowest* of these two approximates the open loop gain $\hat{A}$.

Current gain approximations yield fair results, for instance, for cascaded bipolar transistors in common emitter configuration. The output impedance of each amplifier stage is relatively high compared to the input impedance of the succeeding stage, so that most of the available output current flows into the succeeding stage.

Voltage gain approximations yield fair results, for instance, to simple opamp configurations, in which the input impedance of the opamp is relatively high and the output impedance relatively low.

Note that the current gain of a two-port is quite different from the *open loop* current gain of that two-port. The current gain of a two-port is the ratio between output current and input current, under the condition that the input current is generated using an ideal current source and that the output current is sensed when it is shorted. The same applies for voltage gain and *open loop* voltage gain.

To evaluate the open loop gain from voltage gain and current gain, an additional 'measurement' is required to complete the missing information.

4.3.3. Conclusions

In conclusion, it has been demonstrated that two-cut methods are required to evaluate an exact solution for the loop gain, however one-cut methods provides fair estimations for this exact value. This estimated result is referred as *open loop gain*, and is a characteristic value of two-ports. The open loop gain function is evaluated for various types of two-port parameters.

Further, it has been demonstrated that the open loop gain is independent of the position of the cut. This is a remarkable property because the two-port parameters of the open loop gain two-port change with the position of the cut. It holds for any cascade of two-ports, as long as their cyclic order remains unchanged.

The accuracy of the open loop gain, as approximation for the (exact) loop gain, is higher than that of the open loop *current* gain or the open loop *voltage* gain. These two quantities are representative for the results of commonly used manual loop gain calculations. The relation between all these quantities facilitates a validation assessment of conventional analysis methods.

Two-cut methods are generally preferred, since they provide exact solutions. However, the simplicity of one-cut methods may be decisive in situations with simple bandwidth demands.

4.4. Loop gain representation by poles and zeros

The loop gain transfer function is an important figure of merit in the assessment of the pass band transfer function. The previous sections demonstrated how to determine loop gain, with sufficient accuracy for reliable wideband feedback analysis.

Stability is an important aspect of the pass band transfer function. The determination of loop gain, in terms of poles and zeros, provides valuable information whether a circuit is oscillating or not.

One of the basic topics of this study is the development of a simple but robust pole-zero extraction algorithm. This iterative curve fitting technique fits rational transfer functions to tabular data. The development of this algorithm was essential to the synthesis of stable wideband feedback loops using pole-zero patterns and root-locus techniques.

This section describes how to extract the poles and zeros of a loop with this algorithm, and demonstrates that higher order pole-zero models are reliable description methods for loop gain transfer functions.

4.4.1. The purpose of pole-zero patterns

Loop gain and system gain are transfer functions that can be determined by measurement or simulation. Network analyzers and tabular³ circuit simulators generate frequency tables with magnitude and phase information at distinct frequencies of interest for specifying transfer functions.

The graphical presentation of these tables using Bode plots is complete, however stability performance is difficult to interpret. This is because the difference between stable and unstable performance is implied in a subtle difference in the phase plot of a transfer function. Transfer *measurements* on actual systems provide a direct indication of oscillation since the measurement fails in these situations. Transfer *simulation* may result in innocent-looking Bode plots, lacking any warning of oscillation.

Resolving this uncertainty requires the use of additional presentation methods for tabular transfer functions. A loop gain representation, using pole-zero patterns, is more suitable for stability analysis. This facilitates a reliable assessment of stability because the imaginary axis is the arbitration line between stable poles in the LHP (left half plane) and unstable poles in the RHP (right half plane). The drawback of pole-zero plots is that they provide less information about transfer. Therefore, both Bode plots and pole-zero patterns are required.

4.4.2. Strategy of pole-zero extraction methods

The extraction of poles and zeros from a feedback loop is a delicate and a complex process. Manual as well as automated methods may be used to carry out the extraction *simultaneously* or *sequentially*⁴ with the evaluation of the loop gain. Several methods are listed below:

³ The addition 'tabular' is used in this text to distinct tabular simulators from analytical simulator. A tabular simulator provides a transfer function as sampled data and an analytical simulator as coefficients of functions such as polynomials or rational functions.

⁴ The subdivision in 'simultaneous' and 'sequential' methods is devised in this text.

- *Manual simultaneous* methods use a cascade of first order estimations; each amplifier stage being modeled by one or at most two poles. This approach simplifies the complexity of the overall problem.
- *Manual sequential* methods calculate Bode plots of the loop gain and then estimate its magnitude asymptotes. From these, they estimate the poles and zeros from the location of the corner frequencies. This approach assumes a minimum-phase transfer function that is consistent with all poles and zeros in the LHP.
- *Automated simultaneous* methods translate the complete circuit netlist to a frequency independent matrix format to describe, for instance, the frequency dependent admittance matrix $\mathbf{Y} = \mathbf{Y}_{re} + j\omega \mathbf{Y}_{im}$. The poles and zeros of the transfer function of interest are directly extracted from $\mathbf{Y}_{re}$ and $\mathbf{Y}_{im}$ using eigenvalue calculations. This approach circumvents the evaluation of the complex matrix $\mathbf{Y}$, and is implemented in analytical circuit simulators, such as the programs ANP3 [408,409] and LOCUS [113-118].
- *Automated sequential* methods translate the complete circuit netlist to matrix format, however they repeat the translation for any discrete frequency of interest. The transfer function value, associated with that frequency, is evaluated using matrix inversion calculations. Next, it extracts the poles and zeros using a curve-fit to this transfer table. This approach appears to be cumbersome, however it is wider applicable (measured data, elements with delay) and enables a piece wise evaluation of a complex circuit (significant reduction of round-off errors). It is implemented in tabular circuit simulators, such as the commercial available programs Touchstone® [124], MDS® [126] and many variants of Spice [127,128].

It is out of question that the automated methods are more accurate than the manual methods, however the difference in performance between simultaneous and sequential methods is not obvious.

Analytical circuit simulators suffer more from round-off errors than tabular circuit simulators. This is because simple circuits with element models based on delay may generate a huge number of poles and zeros, while most of them are located far outside the frequency band of interest. A tabular circuit simulator models the same element with a single complex number per frequency and remains applicable when analytical circuit simulators fail. Further, the referred tabular circuit simulators facilitate a piece wise evaluation of a complex circuit, using sub circuits.

As a result, it is plausible that automated sequential methods yield better results than automated simultaneous methods when applied to complex circuits.

One of the results of this study was the development of a simple but robust curve fitting algorithm for the extraction of poles and zeros from tabular transfer functions (see section 3.1.1). Appendix D describes in detail the mathematics of this method. The linear approach, used in this algorithm, combines fast and robust convergence properties and no need to estimate starting values. The proposed algorithm can be implemented simpler than commonly used iteration methods [412,413] such as Levenberg-Marquardt [410] or Gauss-Newton methods [413].

4.4.3. Practical demonstration of the pole-zero extractor

The robustness of the proposed extraction method extends the application of pole-zero analysis to wideband feedback design, since the extracted singularities are representative of actual circuits. This is demonstrated in figure 4.19, in which three realistic examples of lightwave receivers with third order transimpedance feedback are shown.

The basic form of the amplifier stages in figure 4.19a is a common-emitter (CE) stage, to maximize the loop gain. The active decoupling of the stages in figure 4.19b with common-base (CB) stages improves the reverse isolation of the CE-stages because the parasitic local feedback in the preceding transistor is reduced by the low input impedance of the CB-stage. The combined amplifier stages are called *cascode* stages, and cascoding increases the loop gain and improves the smoothness of its magnitude.

The dynamic decoupling of the stages in figure 4.19c facilitates the use of common-collector (CC) stages. These stages are called *long tailed pairs*, and combine the advantages of cascoding with non-inverting transfer, at the cost of a minor reduction of loop gain.

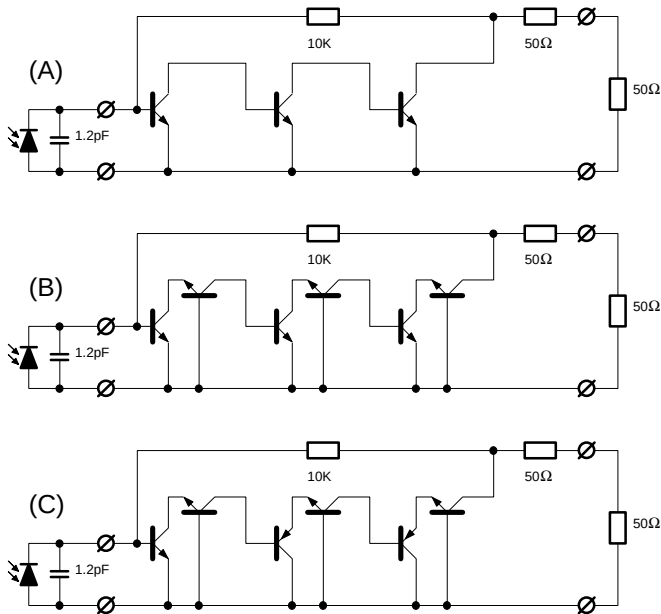


Fig 4.20 Three examples of a transimpedance feedback amplifier with three cascaded amplifier stages: (a) all CE-stages, (b) all cascoded stages and (c) a cascode stage cascaded with two long-tailed pairs.

Figure 4.20 shows the calculated loop gain for these examples. Measured two-port s-parameters of the transistors were used in the simulation to avoid the intermediate step of the extraction of a transistor model. This demonstrated the reliability of the pole-zero analysis. The other circuit elements were modeled.

The same s-parameter data set used here was that used for the BJT transistor model extraction in section 3.2. All transistors are assumed to be equal.

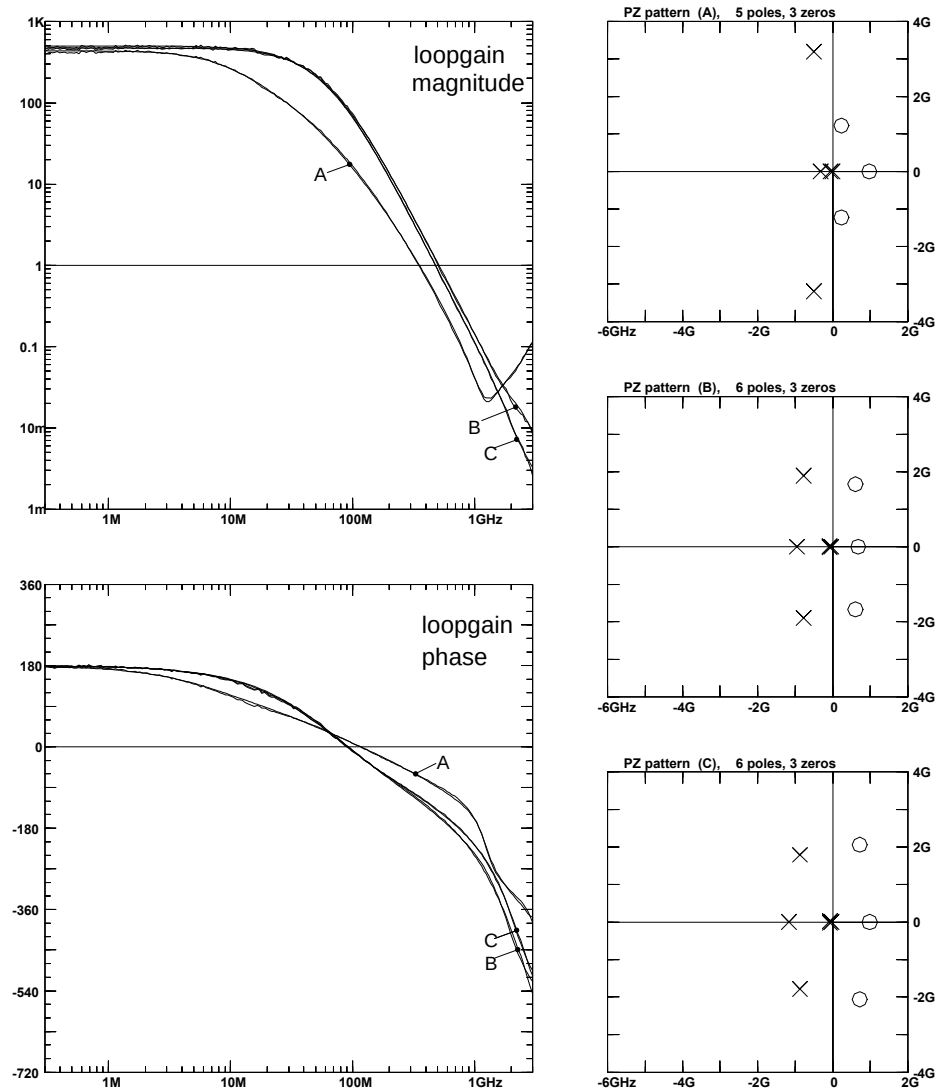


Fig 4.20 Bode plots and pole-zero patterns of the loop gain of the three configurations in figure 4.19. The three data curves in the Bode plots are overlaid with three fitted curves. Note the exceptional fit between model and measured data, while 201 frequency points were used

After the (two-cut) loop gain calculation the simulation fitted three curves of rational functions to the loop gain tables with 201 frequency points each. The choice of the *number* of poles and zeros is a simple process of trial and error, and the algorithm extracts their location. An optimal choice is usually accompanied by a minimum of iteration steps.

The three data curves in the Bode plots of figure 4.20 are overlaid with these fitted curves, and the close match between model and data demonstrates the robustness and the applicability of the method. The associated pole-zero plots in figure 4.20 are therefore accurate representations of the loop gain up to 3 GHz.

One of the most startling observations in figure 4.20 is the presence of RHP-zeros *below* the transition frequency $f_T=3.3$ GHz of the used transistors. These zeros are important obstacles for the synthesis of stable feedback loops, and are discussed in section 5.3. These zeros originate from parasitic effects, and are hardly noticed in Bode plots. They illustrate the importance of reliable pole-zero analysis of loop gain, which will be emphasized in the succeeding sections.

4.4.4. Interpretation of pole-zero patterns

This section is completed with an overview of some interpretation rules for pole-zero patterns.

- Poles and zeros determine the corner frequencies of transfer functions. Above this value, the transfer of poles decreases with the frequency while the transfer of zeros increases. Their magnitude defines the value of the corner frequency and their phase distinguishes between stable, resonant or unstable.
- All LHP-poles indicate a stable transfer function. Poles, closer to the imaginary axis than to the real axis, cause a transfer function with resonance peaks.
- One or more poles located in the right complex half plane of the pole-zero pattern indicate that the circuit is oscillating.
- Zeros are always stable. LHP-zeros are the counterpart of LHP-poles. Pairs of poles and zeros at the same location are a neutral pair.
- RHP-zeros are *not* the counterpart of RHP-poles, because they do not affect the stability of the transfer function. Therefore, these zeros cannot be canceled out by poles.

An alternative is a restricted cancellation by LHP-poles at a mirror position. Mirroring with respect to the imaginary axis causes a complete compensation of the magnitude response, however doubles the phase response.

Figure 4.21 shows some examples of standard pole-zero patterns, with their associated Bode plots.

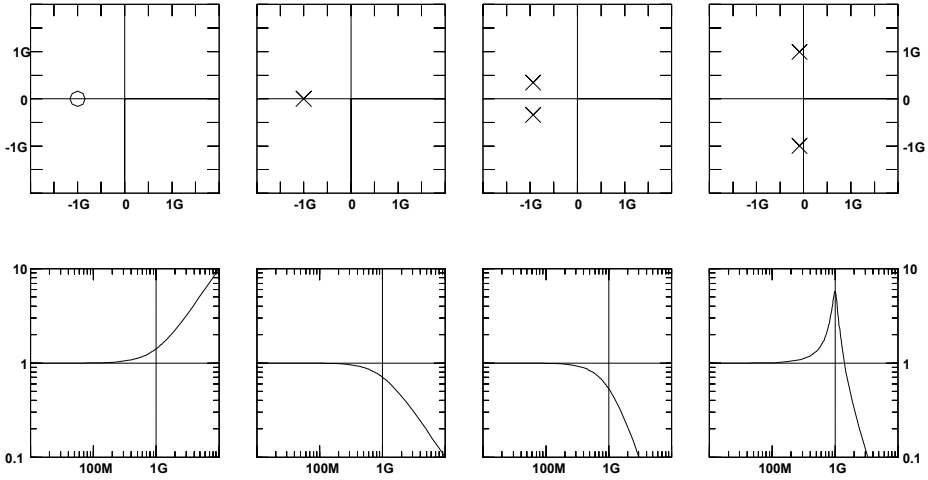


Fig 4.21 Examples of some typical pole or zero patterns, with their associated Bode plot. The axis of the pole-zero patterns are the real and imaginary values of the singularities. The axis of the Bode plots are magnitude versus frequency.

4.4.5. Conclusions

In conclusion, the risks are discussed of feedback analysis that fully relies on Bode plots alone. We emphasized the importance of additional analysis, using pole-zero patterns. Various strategies are implemented to find the poles and zeros of the loop. It has been observed that automated *simultaneous* methods are not very successful in studying complex and wideband circuits. It is plausible that automated *sequential* methods (extraction of poles and zeros, using curve-fitting techniques) yield better results. A simple but robust algorithm for pole-zero extraction is developed, and its robustness is demonstrated with realistic examples. Exceptional good fits, over a very wide frequency band, illustrates the reliability of the proposed method. RHP-zeros were correctly extracted.

The proposed pole-zero analysis does not require any model of the individual transistors. We used *measured* two-port parameters of the individual transistors, combined them in a circuit simulator, and evaluated the (exact) loop gain in tabular form. Pole-zero extraction was the last step of the analysis.

Since an increasing number of manufacturers specify their devices in terms of two-port parameters (a table with measured s-parameters), a powerful tool is developed to perform *reliable* pole-zero analysis, using *realistic* data.

4.5. Loop gain deflation algorithms

Synthesis of stable feedback loops is a delicate process. A *fully* automated procedure for syntheses is currently not available and therefore the assistance of *manual* synthesis remains a substantial part of the overall procedure. The complexity of synthesis requires a simplification of the feedback problem, particularly in the case of manual synthesis.

The major reduction of the complexity of a transfer function is the extraction of poles and zeros from a tabular transfer function. For instance, the example of figure 4.20 has reduced a table with 300 complex numbers to a table with 6 poles and 3 zeros. Yet, any further deflation of the loop gain is desirable to relax the synthesis problem.

Since parasitic effects are essential for wideband feedback design, the commonly used low frequency approach of ignoring them is inadequate.

This study has resulted in powerful reduction algorithms for manual and automated deflation⁵ of loop gain, including:

- A manual deflation method that models all parasitic singularities by *pseudo delay*.
- An automated deflation method that cancels pole-zero pairs that does not coincide *exactly*.
- An automated deflation method that approximates a transfer function with its dominant poles and zeros, within a *predefined* frequency interval.

These reduction techniques are important for the purposes of synthesis, as is described in section 5.3. This section describes the new deflation methods that expand low-frequency synthesis techniques for wideband applications.

4.5.1. Manual deflation using pseudo delay

For the purpose of practical synthesis, it is convenient to distinguish the poles and zeros in two groups: the *primary*⁶ singularities and the *parasitic* singularities. This facilitates a reduction of the system order, the *dominant* system order. The combined influence of all parasitic singularities is then represented by a single real scalar: *pseudo delay*. The here proposed deflation method is mainly intended for manual interpretation of higher-order effects in pole-zero patterns.

The choice whether a singularity is primary or parasitic is arbitrary, and is mainly based on the chosen frequency band of interest. By approximation, poles and zeros have unity magnitude responses for frequencies below half the corner frequency $\omega < \frac{1}{2} |\rho|$, thus all singularities with corner frequencies above twice the highest frequency of interest are good candidates to be classified as parasitic.

Ignorance of all parasitic singularities, which is commonly used in low-frequency design, is not suitable. It was the inspiration of this study to recognize that group delay or phase delay facilitates a fair estimation of the combined parasitic phase effects. As a result, the system order of the loop gain is reduced by the following approximation:

$$\hat{H}(s) \approx \hat{h}(s) \cdot \exp(-s\tau_p) \quad \text{within the basic frequency band}$$

⁵ The word 'deflation' has been adopted from root finding algorithms for polynomials.

⁶ The subdivision in 'primary' and 'parasitic' singularities is devised in this text.

The phase delay of a real zero (Z) or a conjugate pair of complex zeros (Z, Z') at zero frequency ($\omega \rightarrow 0$) is derived as follows:

$$H_1(j\omega) = (1 - j\omega/Z) \Rightarrow \phi \approx \tan(\phi) = -\omega/Z \Rightarrow -\left(\frac{\phi}{\omega}\right) \approx \frac{1}{Z}$$

$$H_2(j\omega) = (1 - j\omega/Z)(1 - j\omega/Z') \Rightarrow \phi \approx \tan(\phi) = -\omega(1/Z + 1/Z') \Rightarrow -\left(\frac{\phi}{\omega}\right) \approx \frac{1}{Z} + \frac{1}{Z'}$$

The same applies for poles, however with opposite sign, and therefore we define the overall pseudo delay as follows:

$$\tau_p = -\left(\frac{\phi}{\omega}\right) = \left[\sum 1/Z_k - \sum 1/P_k \right]_{\text{parasitic}}$$

The magnitude response of a transfer function with pure delay is frequency independent, however this does not apply for poles and zeros. This effect limits the bandwidth of the pseudo delay approximation.

Since poles and zeros have a complementary magnitude response, any mirrored pole-zero pair with respect to the imaginary axis forms an all-pass transfer function. These combinations expand the usable bandwidth of the pseudo delay approximation. When the pairs of poles and zeros lie in Bessel position, then the combined phase delay has a maximally flat frequency response over a wide frequency band.

As is demonstrated in figure 4.22 the same applies for the group delay, however with smaller bandwidth.

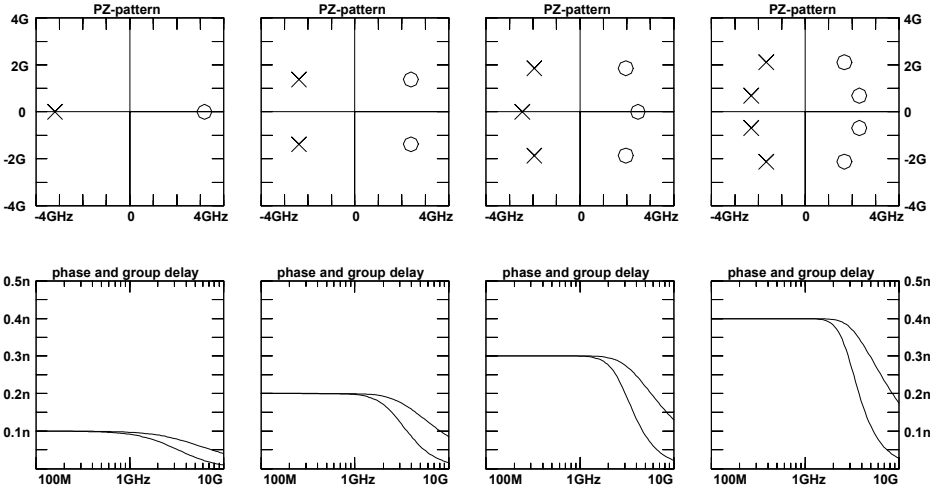


Fig 4.22 Pole-zero patterns and group delay plots of various all-pass filters in which the poles are located in Bessel position. The transfer of these filters is characterized by maximally flat group delay.

The most notable singularities of the specific examples in figure 4.20 are the RHP-zeros. In the examples 4.20b and 4.20c these zeros are associated with poles closely located to the mirror position. As a result, the replacement of these singularities by pseudo delay is a good approximation over a wide frequency band. When all singularities above 500

MHz are classified as parasitic singularities, then the associated parasitic pseudo delay has the following value:

A:	$t_p = 0.23$ [ns]	3 transistors
B:	$t_p = 0.52$ [ns]	6 transistors, all cascoded
C:	$t_p = 0.41$ [ns]	6 transistors, cascode and two long-tailed pairs

The deflation method with pseudo delay is mainly intended for manual synthesis. It is therefore handy to estimate this delay directly from the transistor parameters (see section 3.2.2), without the effort of evaluating loop gain.

As a rule of the thumb, the overall pseudo delay of a cascade of transistors is close to the sum of the delay in all the individual transistors. Cascoding is associated with the highest pseudo delay.

To indicate the importance of these delay figures, a delay of 0.3 ns is equivalent to 9 cm transition time of EM-waves in free space and 4.5 cm in standard epoxy print.

4.5.2. Automated deflation by pole-zero cancellation

Poles are canceled out by zeros that are located at the same position. In that special case, the deflation is achieved by simply removing these pairs. When their positions do not coincide but match within a predefined accuracy, e.g. 10%, then simple removal does not conform to the best feasible approximation. An improved deflation technique must therefore combine the removal of matching pairs with the adjustment of the remaining poles and zeros.

This study has resulted in a powerful deflation algorithm that uses a generalized definition for polynomial division. This subsection describes the application of weighted⁷ polynomial division for reduction of the transfer order.

Pole-zero cancellation is nothing more than division on the numerator $T(s)$ as well as the denominator $N(s)$ of a rational transfer function $H(s)=T(s)/N(s)$ using a common divisor polynomial $X(s)$. An additional complication arises when the quotient is not divisible, which is similar to the round-off error problem in root finding algorithms. As a result, the deflation resolves itself to the following approximation:

$$H(s) = \frac{T(s)}{N(s)} \approx \frac{T_0(s) \cdot X(s)}{N_0(s) \cdot X(s)} = \frac{T_0(s)}{N_0(s)}$$

Various root finding algorithms, as described in section 9.5 of Numerical Recipes [413], divide the calculated roots into the original polynomial. They reduce round-off errors by both forward and backward deflation, followed by polynomial polishing⁸ [413]. A deflation step uses a polynomial division with remainder according to one of these definitions:

$$\begin{array}{ll} T(s) = q_{1r}(s) \cdot X(s) + r_{1r}(s) & \text{normal (backward) division with remainder} \\ T(s) = q_{2f}(s) \cdot X(s) + r_{2f}(s) \cdot s^{m-k} & \text{reverse (forward) division with remainder} \\ T(s) = q_{0g}(s) \cdot X(s) + r_{0g}(s) \cdot W(s) & \text{weighted division with remainder} \end{array}$$

Appendix G describes an algorithm for our weighted division. It can also be used for a conventional polynomial division, although classic methods are simpler.

⁷ The words 'weighed division' are devised in this text.

⁸ Root polishing is an algorithm that reduces round-off errors in polynomials when its roots are divided out.

A good fit for high frequencies arises when normal division is applied and the remainder polynomial (backward deflation) is ignored. Reverse division results in a good fit for low frequencies (forward deflation). To minimize the impact of increasing errors of successively deflated polynomials, some root finding algorithms polish the deflated result. One solution is the combination of the first coefficients in a forward deflation with the last coefficients in a backward deflation. This results in a fair fit for both low and high frequencies, but concentrates the error in the midband region. Study of various deflation results has shown that this concentrated error may cause unreliable results.

A change of the number of coefficients that are taken from the forward quotient polynomial may spread these errors out over a wider frequency band. This type of manual correction improves the result but prevents a full automated solution.

The algorithm proposed here performs deflation and polishing simultaneously by a weighted division and a proper choice for the weighing polynomial $W(s)$. This approach smoothes the error and spreads it out over a wider frequency band.

The relative error $\delta(s)$ of the deflated result is as follows:

$$\delta(s) = \frac{r(s)}{X(s)} \cdot \frac{W(s)}{q(s)}$$

$$\frac{T(s)}{X(s)} = q(s) \cdot (1 + \delta(s))$$

An algorithm for the optimal weighing polynomial $W(s)$ was not found, however a significant improvement was observed when $W(s)/q(s)$ approaches the value one for all midband and highband frequencies and goes to zero for low frequencies. Since the order of $r(s)$ is lower than that of $X(s)$, the proposed choice causes $\delta(s)$ to be zero for zero and infinite frequencies.

The algorithm proposed here starts with an estimation of the quotient $q(s)$ by a conventional mix of forward and backward deflation. It then improves this quotient with an iterative loop of weighted polynomial divisions, using $W(s)=q(s)-q(0)$ as weighing polynomial. Each iteration step performs a successive improvement of the approximation for $q(s)$, and the iteration continues until the variation in $q(s)$ has lost its significance.

4.5.3. Automated deflation using dominant singularities

The concept of dominant singularities is widely used, however its definition is merely intuitive. In low frequency design the concept of *dominant* singularities is equivalent to the concept of *primary* singularities, as defined in section 4.5.1. This approach is not suitable for automated synthesis because a computer aided synthesis procedure requires an unambiguous definition.

This study has resulted in a more suitable definition for dominant singularities that facilitates the reliable prediction of loop gain, using dominant approximations. The definition is just the result of the algorithm that is described in appendix *H*.

Dominant poles and zeros are adjusted primary singularities, modified so that the approximation error of the deflated transfer goes to zero at *both* ends of a predefined frequency interval. The algorithm performs this in two steps:

1. Transform the poles and zeros into polynomial form, in increasing order. Remove the higher order power terms contributing less than a predefined amount, e.g. 30%.

The roots of this reduced polynomial differ from the primary poles and zeros, and facilitate a deflation with better accuracy.

- Polish this intermediate result by adjustment of the two highest polynomial coefficients. This forces the error to zero at the highest frequency of interest.

This polishing step is crucial for the automated compensation synthesis. It facilitates an accuracy improvement at a frequency of choice. The elaboration of the algorithm is described in appendix *H*.

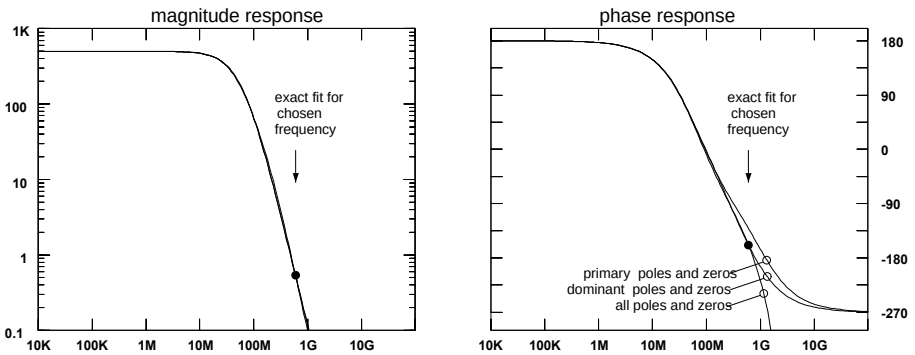


Fig 4.23 Deflation example of a transfer function by primary poles and zeros and by dominant poles and zeros. The polishing step of the dominant deflation caused an exact fit at the chosen frequency (marked). In this example, the magnitude accuracy of both methods is comparative, however the phase accuracy of a dominant approximation is significantly better.

Figure 4.23 demonstrates what can be achieved with the improved method. The transfer function of the example in figure 4.20c was deflated from 6 poles and 3 zeros into 4 poles and 1 zero with different methods. The primary singularities were obtained by discarding the poles and zeros with the highest magnitude. The dominant singularities were obtained by this method.

In this example, the improvement of the phase plot demonstrates the true dominant deflation method is applicable over a wider frequency interval than quasi dominant methods based on primary singularities.

4.5.4. Automated overall deflation

A full automated deflation mixes methods of pole-zero cancellation and dominant approximation. This comprises a cancellation step, followed by a dominant approximation, and completed by a second cancellation step to eliminate pole-zero pairs of intermediate results.

4.5.5. Conclusions

In conclusion, various deflation algorithms are developed for reducing the number of poles and zeros representing the loop gain transfer function. These reduction techniques

are important for the purpose of synthesis. The more reduction of complexity the simpler the synthesis will be, assuming adequate accuracy in a chosen frequency interval.

- A manual deflation method is described to represent all parasitic singularities by a single number: pseudo delay. This provides an attractive figure of merit in estimating an upper limit for the available amplifier bandwidth. Compensation of pure delay is impossible because it would require non-causal compensation networks.
- An automated deflation method is described to cancel pairs of poles and zeros. When they coincide, then a simple removal of these pairs is adequate. Difficulties arise when they match within a predefined accuracy. The proposed algorithm spreads the deflation error out over a wide frequency interval, to find the most plausible solution.
- An automated deflation method is described to replace sets of poles and zeros by sets of dominant poles and zeros. The algorithm provides an exact description of the original transfer function at both ends of a chosen frequency interval. The deflation error is spread the out over the inner frequency interval, while the outer frequency interval is not considered. The performance of this algorithm is superior to simply removing all singularities that exceed a chosen frequency limit.

The proposed algorithms are robust and capable of an adequate reduction of the transfer functions in figure 4.20.

Chapter 5

Wideband feedback synthesis

A reliable feedback analysis is a precondition for wideband feedback design. It facilitates assessing transfer functions of feedback amplifier when their configuration is known. On the other hand, feedback design also requires adequate synthesis. This is the reverse problem, dealing with the question what circuit configurations are required to realize transfer functions that meet predefined requirements.

State of the art / manual methods

Nordholt [406] has developed a manual synthesis method for negative feedback amplifiers. This hierarchical design method was initially intended for feedback amplifier design, however it is wider applicable. This synthesis is roughly characterized in the following design steps:

1. *Basic design*, using simplified models for the individual circuit elements. In this design stage, the principle limitations of the overall circuit performance are analyzed, such as bandwidth and noise.
2. *Analysis*, using manual methods or circuit simulation. In this design stage, the analysis of the circuit is complete, using models that take various parasitic effects into account. The parasitic effects that are ignored in the first design step will degrade the overall performance.
3. *Active decoupling*. Suppressing of parasitic effects, using additional sub-circuits. In this design stage, the amplifier stages are supplemented with buffers, cascoding configurations, etcetera, to minimize degradation by parasitic effects. The goal is to approximate the results of the basic design. (stage 1).
4. Sometimes additional sub-sub-circuits are required, to supplement the sub-circuits of the previous design steps.

Essential for this strategy is that the designer is aware of the degradation from optimal feasible performance, at each stage of the design. Furthermore, the performance of the sub-circuits (and sub-sub-circuits) makes lower demands then the performance of the basic design.

State of the art / automated methods

The above synthesis was initially intended as an overall manual method, however, circuit simulation¹ has expanded the applicability. Circuit simulators automate the analysis, but not the synthesis. Attempts have been made to incorporate an automated synthesis. Recent examples of synthesis programs are [501], IDAC [502], BLADES [503],

¹ Examples of commercially available simulators with powerful N-port facilities are TOUCHSTONE/LIBRA, MDS and Super-Compact/Harmonica. Currently, many others simulators are available with adequate microwave facilities. Further many variants of Berkeley SPICE are commercially available. Touchstone and Libra are courtesy of EEsof inc., MDS of Hewlett Packard and Super-Compact and Harmonica of Compact Software inc.

OASYN [504] OPASYN [505], SEAS [506], OAC [507], ODIN [508], AMPDES [510] and ARIADNE [512].

Due to the complexity of the synthesis problem, the above synthesis programs provide restricted solutions. They may restrict [510] the technology (e.g. CMOS), restrict the number of circuit configurations, use overly simplified models, or rely on adequate reduction of parasitic effects. As a result, they are not applicable to wideband feedback design.

The same applies for the manual use of the above mentioned synthesis theory. The first two design steps are applicable to wideband feedback design. The third step, that must cancel out parasitic effects with additional subcircuits, will eventually fail for increasing bandwidth demands. As a rule of the thumb, above 10% of f_T for second order loops and above 2% of f_T for third order loops, parasitic effects becomes crucial. The sub-circuits will then add additional parasitic effects that may degrade the overall performance instead of improving it.

Wideband feedback design is characterized by parasitic effects that are dominant. So far, there is in general no answer to the question of dealing with parasitic effects that modifies the dominant behavior of the feedback loop.

A novel approach

This chapter provides novel mathematical tools, to synthesize the stability compensation of the loop. It relies on the generalized feedback analysis of the previous chapter 4. that takes fully accounts for parasitic effects.

The proposed synthesis method is partly manual and partly automated. A (tabular) circuit simulator, with augmented facilities, is required to extract the loop gain using the theory of chapter 4. The required transfer functions of the compensation networks in the loop are synthesized with a novel compensation algorithm, that is discussed in this chapter 5. Additional (conventional) synthesis remains required to translate the synthesized compensation transfer into practical circuits.

The highlights of this chapter, which are developed in this study, are:

- Development of a generalized theory on passive compensation networks, that is suitable for automated compensation synthesis.
- Development of an automated synthesis algorithm that predicts the required compensation, in terms of poles and zeros. It has found solutions that are commonly excluded from being feasible (see example in section 5.3.4).

5.1. Aperture analysis of feedback amplifiers

When the loop $\hat{H}(s) = A(s)\beta(s)$ is closed, the system gain is controlled by feedback. In general, the loop requires special compensation measures to avoid a feedback system that is unstable or that has resonant peaks in its transfer function.

Compensation syntheses requires an adequate definition for the design goal, and an adequate estimation of the design limitation. The ratio between desired system gain and realized system gain is an intuitive measure to study the pass band limitations of the system. This text refers to this pass band transfer function with the new term *aperture*.

Defining the design goal is equivalent with specifying the aperture. The desired (generic) passband is selected from a class of low-pass filters, selected by the designer. This normalized transfer function is subsequently scaled to its actual value, using the available bandwidth as scaling factor.

This section analyses the aperture in terms of available bandwidth and suitable normalization of the bandpass transfer function.

At first, the concept *aperture* is defined. Some textbooks use the term *closed loop gain* for what will be defined here as "virtual-aperture" while other textbooks use this term for what is defined here as "realized system gain". To avoid further confusion, this study has made up the new term *aperture*.

This study has developed a formal description for the design goal of optimal feedback loops. This formal approach has forced to make a clear distinction between the commonly used *virtual-aperture* and the here proposed *effective-aperture*.

Further, this study has resulted in reliable methods to calculate the available bandwidth. These methods are robust and preserve their validation for feedback loops with parasitic poles and zeros. Finally, this study has resulted in a unified method to specify the generic transfer that is realizable for the passband.

5.1.1. Definition of virtual-aperture and effective-aperture

In general, a loop without any stability compensation will oscillate or will have resonant peaks. The aim of compensating networks is the improvement of the system gain. Since the desired system gain is not necessary a constant, as in case of an integrator or an equalizer, the direct use of system gain is not a convenient design goal. A general applicable design goal is the ratio between what was intended and what is realized.

This text uses the name *aperture* for transfer functions that describe the ratio between the system gain that is desired and the system gain that is realized. In figurative language: the aperture separates the desired gain from the realized gain by a low-pass filter that limits the signal bandwidth when it passes this virtual 'opening'. The design goal is an aperture value close to one, over the widest possible frequency band.

A formal analysis requires an unambiguous definition of desired system gain; a definition that is invariant for any change in compensation. When (β, ϵ, ν) refer to the parameters of the compensated feedback network and $(\beta_0, \epsilon_0, \nu_0)$ to the same parameters of the uncompensated network, then the desired system gain is defined as follows.

<i>forward leakage:</i>	$A_{t0} = \rho$
<i>realized system gain:</i>	$A_t = -(\varepsilon \cdot v / \beta) \cdot (A\beta) / (A\beta - 1) + \rho$
<i>asymptotic system gain:</i>	$A_{t\infty} = -(\varepsilon \cdot v / \beta) + \rho$
<i>virtual system gain:</i>	$A_{tv} = -(\varepsilon \cdot v / \beta)$
<i>desired system gain:</i>	$A_{tt} = -(\varepsilon_0 \cdot v_0 / \beta_0)$

Our definition of desired system gain is valid when the forward leakage (ρ) is small, compared to the system gain. Furthermore, we assume that an unambiguous distinction can be made between compensated and uncompensated state. This distinction is not always evident and has led us to introduce in this text two aperture definitions.

We define *effective-aperture* that is of general applicability, and *virtual-aperture* that is of limited applicability. They are defined as follows:

Effective-aperture	$= \frac{\text{RealizedGain} - \text{Leakage}}{\text{DesiredGain}}$	$= \frac{\hat{H}}{\hat{H} - 1} \cdot \frac{\varepsilon \cdot v \cdot \beta_0}{\varepsilon_0 \cdot v_0 \cdot \beta}$
Virtual-aperture	$= \frac{\text{RealizedGain} - \text{Leakage}}{\text{VirtualGain}}$	$= \frac{\hat{H}}{\hat{H} - 1}$

$\hat{H}(s) = A(s) \cdot \beta(s)$

Compensation is used to realize a system gain that approximates the *desired* system gain. The effective-aperture is a handy figure of merit for studying the compensation. An adequate compensation forces the effective-aperture value to one within the frequency band of interest, without resonating peaks and oscillations.

When all compensation is concentrated in the forward amplifier, the virtual-aperture is equal to the effective-aperture. A distinction between those two apertures is then irrelevant. In all other cases the virtual-aperture is a misleading measure for assessing the compensation. Nevertheless, it is a commonly used measure, since the loop gain is the only transfer function that is required for its evaluation.

In order to demonstrate the use of effective-aperture and virtual-aperture in practice, it is convenient to model the compensation as though it is concentrated in two distinct networks. Since the compensation is most effective in $A(s)$ and $\beta(s)$, this example assumes that the compensation does not affect $\varepsilon(s)$ and $v(s)$. The loop gain is then split-up in the following basic elements:

<i>uncompensated forward gain</i>	$A_0(s)$
<i>uncompensated feedback factor</i>	$\beta_0(s)$
<i>profiled compensation</i>	$H_{cs}(s)$ cascaded with $A_0(s)$
<i>phantom compensation</i>	$H_{cp}(s)$ cascaded with $\beta_0(s)$

The application of the words *profiled* and *phantom* is clarified in a succeeding section 5.1.2. In this section these names are merely used to distinguish the compensation transfer in the forward amplifier (profiled) and the compensation transfer in the feedback network (phantom).

Figure 5.1 shows two sets of equivalent flow diagrams that represent the system feedback in the case (5.1a) that the compensation in the loop is made explicitly visible and in the case (5.1b) that the compensation is embedded. In the first example the

desired system gain and the associated effective-aperture are in evidence. In the second example the virtual gain and the virtual-aperture are used, since no information is available on the phantom compensation.

This figure illustrates that the main signal path is split into two parts: a gain block with ideal characteristics and a (low-pass) filter block that represents all bandpass limitations. It demonstrates that the effective-aperture meets the requirements of the design goal, for which the bandwidth must be maximized and the passband equalized. In this example, the effective-aperture reduces to:

$$\text{Effective-aperture} = \frac{\hat{H} / H_{cp}}{\hat{H} - 1} = \frac{1/H_{cp}}{1 - 1/\hat{H}}$$

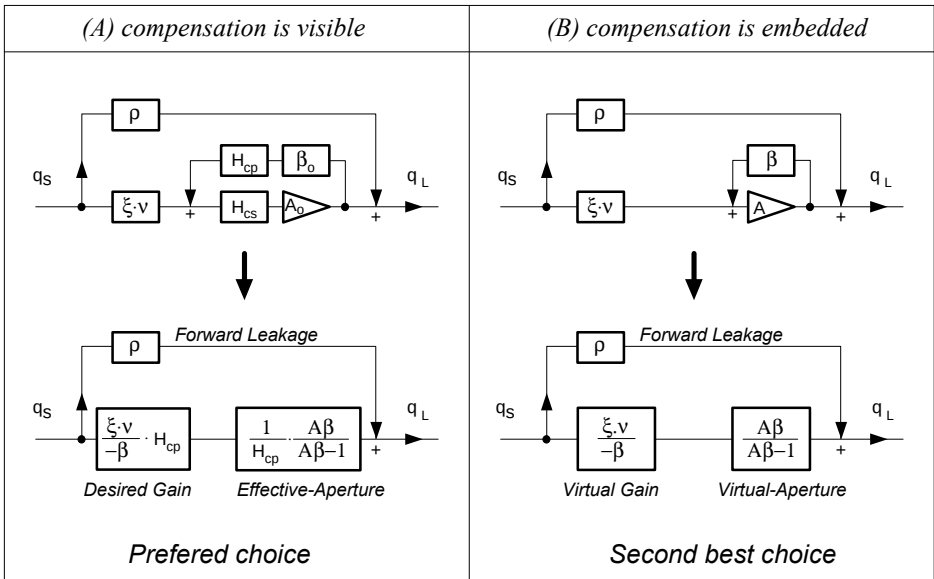


Fig 5.11 Distinction between effective-aperture and virtual-aperture. The aperture is the transfer of a virtual low pass filter that limits the bandwidth of the gained signal. The effective-aperture is a measure that remains valid when the compensation is (partly) phantom ($H_{cp} \neq 1$). The virtual-aperture is often used, however will be misleading for bandpass analysis. In other situations, virtual-aperture and effective-aperture are equal.

5.1.2. Bandwidth analysis

The estimation of the available bandwidth of the effective-aperture has been a standard design aspect from the very beginning of wideband amplifier design. It serves as an initial design goal and may be used to distinguish efficient from inefficient compensation methods. This has resulted in the well-known gain-bandwidth product (*GB-product*) for first order feedback loops and, more recently [406], in the loop gain-poles product (*LP-product*) for all-pole higher order feedback loops.

In its current form, the LP-product is not suitable to predict the available bandwidth for loops with annoying parasitic poles and zeros. This is because the LP-product is restricted to *dominant* singularities, that cannot be extracted when the available bandwidth is unknown. As a result, a complementary algorithm is required.

This study has resulted in a robust algorithm that predicts the available bandwidth, irrespective of the presence of parasitic poles and zeros.

For reasons of completeness, the intuitively based design process commonly followed are reviewed here. The aim of this subsection is, however, to develop a method for formal analysis as required by automated synthesis.

Definitions

To begin, the concept *available* is introduced in the context of asymptotic approximations of the effective-aperture. In a well-designed feedback amplifier, the effective-aperture is close to one, over the widest possible frequency band.

The compensation affects this bandwidth in such a way that passive compensation degrades the bandwidth while active compensation improves it. This is because active compensation networks are amplifiers that provide additional loop gain, and an overall increase in loop gain may result in increased bandwidth. Passive compensators, on the other hand, are attenuators that reduce the loop gain and possibly the bandwidth.

This chapter is restricted to passive compensation. These compensation networks are optimal when they preserve all available high-frequency gain. This facilitates the realization of amplifiers with the widest bandwidth.

In a well-designed feedback amplifier, the effective-aperture is constant over a wide frequency band, tending to zero at very high frequencies. Since stability problems are mainly reserved for mid-band frequencies, the compensation must restrict itself to midband frequencies, and must leave the effective-aperture unchanged for zero and infinite frequencies. This facilitates a simplified calculation of the asymptotic effective-aperture without further knowledge of the compensation.

Phantom- and profiled compensation influence the effective-aperture in a different way because of their different location in the loop. Various loop gain aspects can therefore be isolated to simplify the asymptotic evaluation for the effective-aperture:

<i>profiled compensation:</i>	$H_{cs}(s)$	$g_s = H_{cs}(0)$
<i>phantom compensation:</i>	$H_{cp}(s)$	$g_p = H_{cp}(0)$
<i>uncompensated loop gain:</i>	$H_0(s) = A_0(s) \cdot \beta_0(s)$	$g_0 = H_0(0)$
<i>profiled loop gain:</i>	$H_s(s) = A_0(s) \cdot \beta_0(s) \cdot H_{cs}(s)$	$g_s = H_s(0)$
<i>total loop gain:</i>	$\hat{H}(s) = A_0(s) \cdot \beta_0(s) \cdot H_{cs}(s) \cdot H_{cp}(s)$	$g = \hat{H}(0)$

Figure 5.2 illustrates the difference between profiled loop gain and total loop gain for an example of a compensated amplifier with three dominant poles. Two (phantom) zeros in the feedback network are added to the loop, to provide the overall compensation. They

modify the loop gain at high frequencies, which causes the difference between profiled and total loop gain.

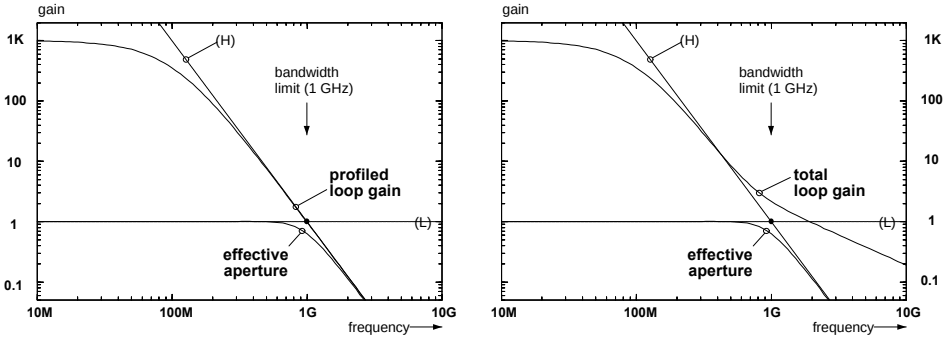


Fig 5.2 Difference between profiled loop gain and total loop gain for a third order loop with two phantom zeros. The point of intersection of the two asymptotic lines (H) and (L) is the bandwidth limit (1 GHz) of the effective-aperture.

The profiled loop gain predicts the asymptotic line (H), and is a good measure for the available bandwidth. This does not hold for the total loop gain.

Asymptotic approximations

In a well-designed compensated amplifier, the dominant order of the profiled part of the loop gain is higher than the dominant order of the phantom compensation. This property facilitates an asymptotic approximation of the effective-aperture for $\omega \rightarrow \infty$. Neglecting the high frequency transfer of the phantom compensation, with respect to the profiled part of the loop gain, results in:

$$A_{sa}(s) = \frac{1/H_{cp}(s)}{1 - 1/\dot{H}(s)} = \frac{1}{H_{cp}(s) - H_{cp}/\dot{H}(s)(s)} = \frac{1}{H_{cp}(s) - 1/H_s(s)}$$

$$\Rightarrow \left\{ \begin{array}{ll} A_{sa}(s) \rightarrow -H_s(s) & (\text{for } \omega \rightarrow \infty) \\ A_{sa}(s) \rightarrow \frac{1}{H_{cp}(0)} = \frac{1}{g_p} & (\text{for } \omega \rightarrow 0) \\ |A_{sa}(j\omega)| \approx \frac{1}{|g_p| + 1/|H_s(j\omega)|} & \end{array} \right.$$

The above approximations uses the property that the profiled loop gain is significantly higher than one for low frequencies and tends to zero for high frequencies. As a result, the transfer of the effective-aperture tends to $(1/g_p)$ in the pass band, and tends to $-H_s(s)$ for higher frequencies.

Figure 5.2 demonstrates the impact of this approximation. The close match between effective-aperture and its asymptotic approximation facilitates a first estimation of the

available bandwidth on the basis of asymptotic lines. Note that the high frequency asymptotic line is not an approximation of the *total* loop gain but of a stripped variant, the *profiled loop gain*.

The generalized bandwidth algorithm, in two steps

Secondly, a two-step procedure is described for determining the available bandwidth, that uses the asymptotic properties of the effective-aperture. The first step is a rough estimation, based on the asymptotic corner frequency of the profiled loop gain magnitude. The second step is a refinement step that uses this estimation to deflate the loop gain to its dominant form, using the reduction techniques of section 4.5. The asymptotic corner frequency of the dominant form is then determined, this being a generalization of the LP-product.

Manual synthesis can not use the improved accuracy of the second refinement step due to the approximate nature of manual methods. When manual synthesis methods are adequate, the two-step approach does not result in significantly improved results.

We have observed that the here proposed refinement step is crucial for automated synthesis in those cases in which manual methods fail due to the complexity of the feedback loop. This is because the available bandwidth estimation is an important input parameter for the automated synthesis algorithm that is described in section 5.3. The refinement step has enabled the algorithm to predict realistic solutions where it would otherwise have failed. The generalized definition of the dominant loop gain, see section 4.5.3, is essential to the success of this refinement step.

Step 1: Global asymptotic corner frequency. A rough estimation of the available bandwidth is the global asymptotic corner frequency. This frequency ω_b is defined as the intersection point ω_b between low frequency approximation $|A_{sa}(0)| \approx 1/|g_p|$ and the high frequency approximation $|A_{sa}(j\omega)| \approx H_s(j\omega)$ of the effective-aperture. Its value is close to the frequency where the product of phantom dc-gain and profiled loop gain has the value one: $|g_p \cdot H_s(j\omega_b)| = 1$.

For manual analysis it is satisfactory to draw a Bode plot of the profiled loop gain and to read off the intersection point with the unity gain line. Automated analysis may follow the same procedure, using a standard iterative root finding method.

From the standpoint of software implementation, it is convenient to re-use the available polynomial root finding algorithm for this purpose. The corner frequency ω_b is then evaluated as below by the calculation of polynomial roots.

$$|g_p \cdot H_s(j\omega_b)| = |g_p \cdot T_s(j\omega_b)/N_s(j\omega_b)| = 1 \quad \text{is a rational function}$$

$$|g_p \cdot H_s(j\omega_b)|^2 = |g_p|^2 \cdot H_s(+j\omega_b) \cdot H_s(-j\omega_b) = 1$$

$$|g_p|^2 \cdot T_s(+j\omega_b) \cdot T_s(-j\omega_b) - N_s(+j\omega_b) \cdot N_s(-j\omega_b) = 0$$

$$\boxed{\{j\omega_b\} \approx \text{roots}(|g_p|^2 \cdot T_s(+j\omega_b) \cdot T_s(-j\omega_b) - N_s(+j\omega_b) \cdot N_s(-j\omega_b))}$$

In general, a set of real, purely imaginary and complex pairs will be found as solution for $\{j\omega_b\}$. The purely (positive) imaginary roots are of interest, and usually there is a

unique solution. In the exceptional case that there exist more than one intersection between profiled loop gain and the unity gain line, the highest intersection is will be the desired corner frequency.

Step 2: Dominant asymptotic corner frequency.

The dominant asymptotic corner frequency ω_0 is an improved estimate of the available bandwidth is . The definition of ω_0 is similar to that of the global asymptotic corner frequencies ω_b differing in the use of a dominant approximation of the profiled loop gain. It requires the preceding calculation of the global value ω_b to supply the corner frequency for the dominant deflation algorithm of section 4.5.3.

From the magnitude approximation of the effective-aperture we obtain:

$$|A_{sa}(j\omega)| \approx \frac{1}{|g_p| + 1/|H_s(j\omega)|} \approx \frac{|A_{sa}(0)|}{1 + |\omega/\omega_0|^{n-m}}$$

$$H_s(s) \approx g_s \cdot \frac{\prod_{r=1}^{r=m} (1-s/z_r)}{\prod_{k=1}^{k=n} (1-s/p_k)} \quad (\text{dominant approximation})$$

$$|A_{sa}(0)| = \frac{1}{|g_p| + 1/|g_s|} = \frac{|g_s|}{|g_p \cdot g_s| + 1} = \frac{|g_s|}{|g_0| + 1}$$

$$|A_{sa}(j\omega)| \approx |H_s(j\omega)| \approx |g_s| \cdot \prod |p_k| / \prod |z_r| \cdot |1/j\omega|^{n-m}$$

$\omega_0 = \sqrt[n]{(1+ g_0) \cdot (\prod p_k) / (\prod z_r)}$	$\eta = (n-m)$

$\{p_k\}$	= all dominant profiled poles
$\{z_r\}$	= all dominant profiled zeros
$g_0 = g_s \cdot g_p$	= total DC loop gain
n, m	= number of dominant profiled loop gain poles and zeros

The total DC loop gain and the dominant poles and zeros of the profiled loop gain are the only parameters that determine the dominant asymptotic corner frequency.

5.1.3. Bandpass analysis

The available bandwidth of the effective-aperture is in essence a frequency scaling factor of all possible generic low-pass responses for the passband. The bandpass transfer function comprises the whole effective-aperture and must be flattened and made free of resonance.

Synthesis algorithms for compensation require a precise definition of the desired bandpass transfer function, such as maximally-flat or equal ripple amplitude, or linear-phase. This section describes some standard transfer functions, adopted and modified from low-pass filter theory, that match the available bandwidth definition of section 5.1.2.

Figure 5.3 shows some examples of standard low-pass all-pole transfer functions for various transfer orders. These plots are normalized with respect to their asymptotic corner frequency, which is defined as the point of intersection between the two asymptotic Bode lines for low and high frequencies.

The asymptotic corner frequencies proposed here differ from the -3 dB corner frequency that is commonly used in filter theory. This is to link better up with the available bandwidth theory of section 5.1.2. The exact definitions of these modified curves and the algorithms for the associated transfer coefficients are developed during this study, and summarized in appendix *I*.

Compensation synthesis starts from the calculation of the available bandwidth and from the selection of an appropriate bandpass curve. The choice of the low-pass curve is an important design parameter, and is application dependent. On account for what grounds must these choices be based?

One of the most important criterion for digital signals is the minimization of overshoot in the step response. For a set of modulated signals, such as TV channels, a magnitude response with maximum bandwidth (-3 dB) might be relevant. With respect to the three standard transfer functions, relevant properties for a given asymptotic corner frequency are summarized below:

- When the low-pass transfer is close to a Bessel transfer function, the group delay characteristic is maximally flat and the step response has no overshoot.
- When the transfer is close to a Butterworth transfer function, the magnitude response is maximally flat and has a wider bandwidth. Note for digital signals that the lack of ripple in the amplitude response is not equivalent to a lack of overshoot in the step response. As a result, the rise and fall times of the Butterworth step response are shorter than for the Bessel response. This benefit is accompanied by some overshoot in the step response (e.g. 10%).
- The bandwidth of a Chebyshev transfer function, with equal asymptotic corner frequency, is even wider. This is, however, at the cost of significantly higher overshoot.

It is quite risky to restrict the passband analysis to frequency domain analysis alone. Using a frequency domain circuit simulator as a design tool, can easily lead to amplifiers with a seemingly excellent magnitude response, that nevertheless exhibits oscillatory behavior. Although the required information is available in a Bode plot, the difference in phase response between a parasitic pole slightly to the left or slightly to the right of the imaginary axis is hardly noticeable. It is therefore necessary to observe the pole-zero patterns of the aperture of the feedback loop, e.g. as a root-locus plot of the loop gain. The pole-zero patterns in figure 5.3 of characteristic transfer functions are useful for pattern recognition.

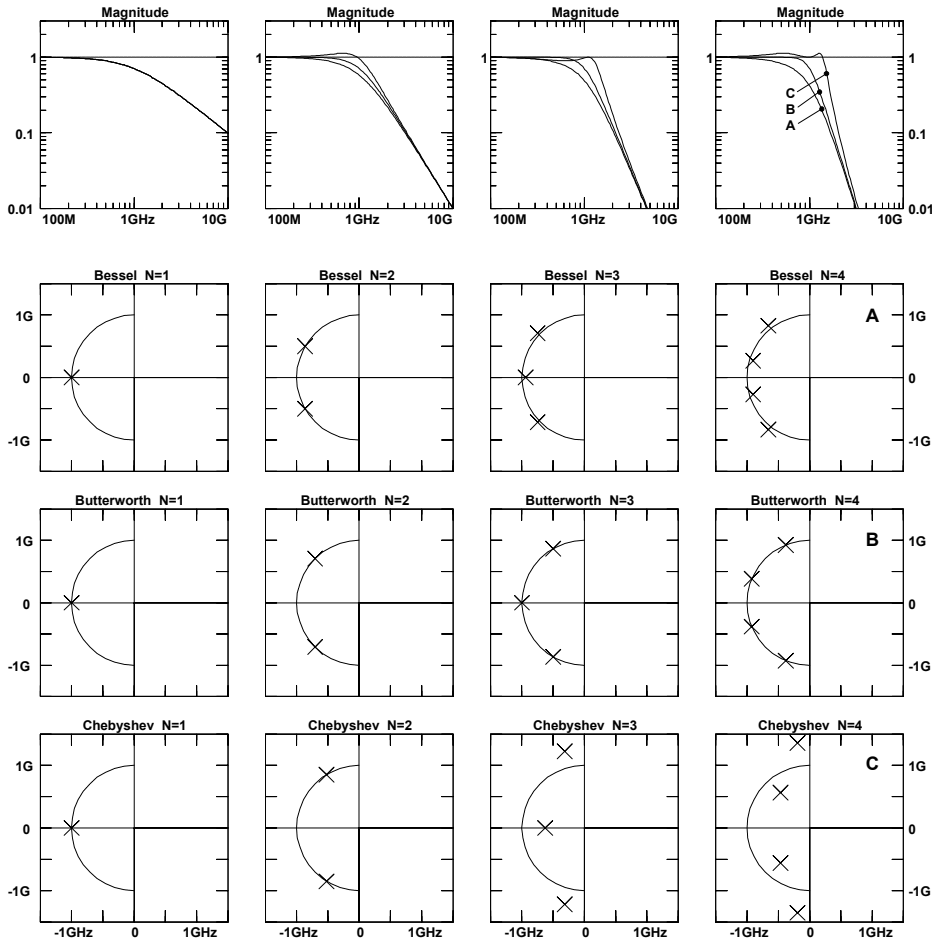


Fig 5.11 Pole-zero patterns and magnitude plots of various standard low-pass filters: Bessel (A), Butterworth (B) and a 1 dB Chebyshev (C) filter.

5.1.4. Conclusions

In conclusion, we defined a transfer to enable a designer to define an adequate design goal for feedback synthesis purposes. This so called *aperture* represents the ratio between desired system gain and realized system gain, and has a low-pass transfer function.

Two different apertures were defined: the *effective-aperture* that is of general applicability and the *virtual-aperture* that is of limited applicability. The use of the effective-aperture is preferred, however it requires the transfer function of the phantom compensator in the loop gain. When this is impossible, the designer must manage a restricted synthesis with the virtual-aperture.

We developed a robust algorithm that estimates the available bandwidth from the asymptotic behavior of the effective-aperture. Since a rough estimation may fail when designing complex feedback loops, an additional refinement step was developed. When future circuit simulators give full support to bandwidth prediction, then the interaction between simulator and designer will be as follows:

- The designer splits the circuit in a forward amplifier and a feedback network.
- The simulator extracts the superposition parameters and all the poles and zeros of the loop.
- The designer indicates what feedback singularities are inserted for phantom compensation.
- The simulator provides the available bandwidth.

When the simulator gives full support to the synthesis of section 5.3, then it also provides the required transfer functions for an (additional) phantom compensator and an (additional) profiled compensator.

The available bandwidth is in essence a frequency scaling factor of all possible generic low-pass responses for the passband of the effective-aperture. Various all-pole low-pass responses are analyzed to enable the designer in defining unambiguous design goals for synthesis purposes. Various selection criteria are discussed for three generic low-pass responses: Chebyshev, Butterworth and Bessel. The concept of asymptotic corner frequency is described for low-pass transfer functions to facilitate an adequate normalization.

5.2. Compensation techniques for feedback amplifiers

Compensation is aimed at modifying the loop gain so that the effective-aperture better approximates a predefined transfer function. Practical compensation is confined to physically realizable networks. A number of successful compensation techniques have been described in the literature, by Ghausi and Pederson [404], Cherry and Hooper [405] and Nordholt [406]. These techniques are analyzed primarily from an implementation point of view.

Benefit of optimizers

Automated optimization techniques are very effective compensation tools when the circuit configuration is chosen. The optimizer adjusts various element values simultaneously in an iterative way, to fulfill a predefined design goal. The commercial available simulators Touchstone® [124], Libra® [125], and MDS® [126] give full support to this approach. They allow a hierarchy of circuit blocks and permit the use of variable labels to define relations between tunable elements.

The aperture analysis of section 5.1 provides reliable predictions of the aperture (available bandwidth and generic passband) and is of valuable use to define that design goal. The more realistic this goal is, the more effective an optimizer will be.

The classic paper by Temes and Calahan in 1967 [411] was one of the earliest to formally advocate the use of iterative optimization in circuit design. Bandler and Chen [412] described in 1988 a detailed overview of various methods and summarized more than 100 references to relevant papers. Most optimizers are implemented for tabular circuit simulators. An implementation for an analytical circuit simulator is mentioned in [509,510].

Restrictions of optimizers

Automated optimization techniques require circuit configurations and tuning elements as input. When a solution exists, when adequate tuning elements with proper starting value are chosen and when parasitic effects are adequately minimized by additional sub-circuits (buffering, cascoding), then optimizers may be very effective.

On the other hand, they will never find a solution when the compensating elements are inadequate. Compensation techniques that are effective below, e.g. $f_T/10$, may fail for wideband feedback amplifiers.

An additional approach

Simulators should inform the designer whether it is *principally* impossible to compensate for instabilities or not. If not, it is of great value to inform the designer what transfer function is required for additional compensation. It might help to find an additional compensation network that causes the required additional transfer.

In particular, an abstraction level is required that facilitates the separation of physical from non-physical solutions. It must be isolated from the implementation. Practical implementations of compensating networks are usually intertwined with the circuitry. Without loss of generality, we will act as if these networks are fully separable from the forward amplifier and the feedback factor. This may result in an unorthodox description of compensation, however it is wider applicable.

Section 5.2 begins with the de-embedding of the compensation networks from the forward amplifier and from the feedback factor. It summarizes basic properties that result from the restriction that the compensation be passive.

Particular attention is paid to the transfer properties, with no more than a weak link to practical circuits. The major part of the implementation problem is left to the reader and to the referred publications.

5.2.1. *Profiled compensation techniques in the forward amplifier*

The application of compensation to forward amplifiers (A) and feedback factors (β), is far more effective than modification of the input and output conversion factors (ϵ and ν). The two forms are identified² as *profiled* and *phantom* compensation respectively (see figure 5.1 and the associated subsection).

The name *profiled* refers to the property that all profiled compensation zeros come through in the effective-aperture at the same location. This property does not hold for the *phantom* compensation zeros.

The transfer function of a physical network has at least as many poles as zeros, however in feedback loops the number of poles is always higher. It means that the loop gain goes to zero for infinity frequencies, since all physical devices (media) will eventually block all signal flow when frequency increases, if necessary up to lightwave frequencies.

The phase shift, associated with these poles, causes the loop to resonate at distinct frequencies or to oscillate. This is particularly true for the frequency band in which the loop gain approaches the value one. The more poles in the passband the more trouble they cause.

The phase shift, associated with a single pole, is insufficient to cause the loop to resonate. Compensation is therefore required for higher order loops. The same applies for lowband pole-zero *pairs*, because they add no phase-shift to the gain at highband frequencies. The location of midband and highband poles and zeros is critical.

Profiled compensation is basically the addition of zeros in the loop to cancel out the annoying poles in the passband. Because the addition of compensating zeros comes in pairs with new poles, the uncompensated pole is replaced in practice by a new pole positioned at a more favorable location. This is only effective when the new pole position is associated with an optimal trade-off between minimal phase shift and optimal preservation of loop gain. As a result, the essence of profiled compensation is to cover (midband) poles with compensation zeros, and then to add (highband) poles that are precisely positioned at a more favorable location.

The exact requirements for the location of these poles and zeros are discussed in section 5.3; this subsection restricts itself to the overall compensation principle.

² The word *profiled* has been introduced in this text. The word *phantom* is commonly used.

Generic magnitude response

Based on the assumption that in the bandwidth analysis all amplification effort is put in the design of the forward amplifier, the profiled compensation will most likely be passive. An optimal passive profiled compensator $H_{cs}(j\omega)$ must conform to the following requirements:

- It preserves all gain at the high-end of the passband to preserve the available bandwidth. This means that $H_{cs}(j\omega) \rightarrow 1$ in the high frequency band.
- It preserves all gain at the low-end of the passband because modification of a frequency band with high gain has nothing to add to the overall stability. This means that $H_{cs}(0) \rightarrow 1$ in the low frequency band

As a result, an optimal profiled compensator is an attenuator of mid-band frequencies only. A near-optimal profiled compensator causes additional attenuation at low-band frequencies. This is illustrated in figure 5.4.

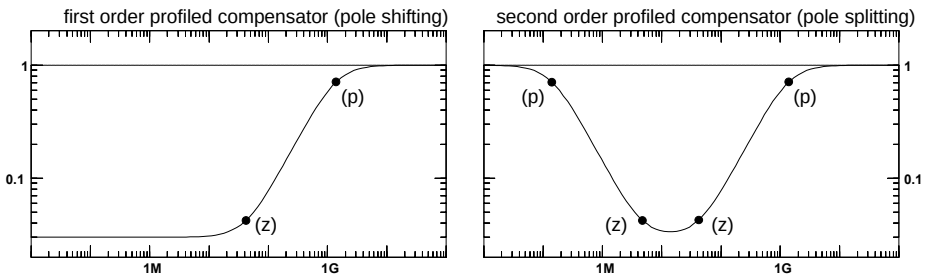


Fig 5.4 Typical example of the magnitude response of a passive profiled compensator that is cascaded with the forward amplifier. Midband frequencies are attenuated to realize the compensating zeros. All high-band frequencies are passed to preserve the available bandwidth. Attenuation of low-band frequencies deteriorates the feedback accuracy, moderately affecting the passband of the effective-aperture.

First order profiled compensators, although providing effective compensation, are unable to preserve the gain for low-band frequencies. At minimum, a second order transfer function, or higher, is required to preserve the gain at both ends of the passband. The higher the transfer order, the better the compensation may be.

Another aspect illustrated in figure 5.4 is that it is a bad policy to position the poles, associated with the compensation, at a corner frequency that is *significantly* higher than the bandwidth limit. This causes a reduction of loop gain at high-band frequencies and with that a reduction of available bandwidth.

Generic pole-zero pattern

The magnitude response in figure 5.4 of a second order profiled compensator is typical for optimal higher order profiled compensators: all-pass functions with a dip at mid-band frequencies. This implies that the number of compensating zeros is the same as the number of poles belonging to them.

Further, the gain is approximately one, for frequencies exceeding the bandwidth limit. This implies that the product of all poles equals the product of all zeros.

The position of the poles and zeros is critical. It happens that the zeros coincide with the poles in the special case that *all* singularities in the loop are compensated. This means that the transfer order of the compensating network equals the transfer order of the uncompensated loop. As a result, the overall system order will not increase when employing full compensation. In the more general case where the compensation is restricted to the *dominant* singularities, the zeros will not necessarily coincide with the poles.

5.2.2. *Profiled compensation techniques with intertwined solutions*

From a mathematical point of view, the profiled compensation is virtually concentrated in a single equalizing network, as though it is cascaded with the forward amplifier. In a practical implementation, the profiled compensation is distributed and embedded in the amplifier stages. The transfer of an uncompensated amplifier stage and its compensation are then intertwined by proper modification of the original transfer function.

There are no fundamental reasons to avoid equalizers as profiled compensators, however from a practical point of view intertwined solutions are more attractive due to their simplicity. Figure 5.5 and 5.6 show examples of intertwined solutions, that are commonly used.

This study has led to increased insight into the principal restriction of intertwined solutions. They modify the uncompensated transfer function, which implies that the uncompensated poles are moved from their original position. This is equivalent to a profiled compensator that matches some compensating zeros with uncompensated poles. As a result, intertwined solutions have fewer degrees of freedom than distinct solutions. An intertwined solution may be sub-optimal, when the presence of parasitic singularities restricts the compensation to a dominant form.

Intertwined compensation by pole shifting

Figure 5.5 shows an example of an intertwined first order profiled compensator. An additional resistor competes with the, mainly capacitive, input impedance of a transistor and absorbs the current gain at low-band frequencies. The asymptotic Bode plots illustrate that this intervention replaces the uncompensated transfer with a mid-band pole by a compensated transfer with a pole at a higher corner frequency. This is called *pole shifting*.

The compensation transfer function is a virtual transfer function that originates from the ratio between compensated and uncompensated current gain. Note that the pole-zero pair of compensation zero and uncompensated pole is available in this virtual transfer function but is absent in the actual current gain.

Practical implementations will always suffer from some degradation of high frequency gain. In this example the addition of a well-adjusted inductor, in series with the absorbing resistor, will relax its absorption at high-band frequencies. Such an inductor will improve the loop gain at higher frequencies.

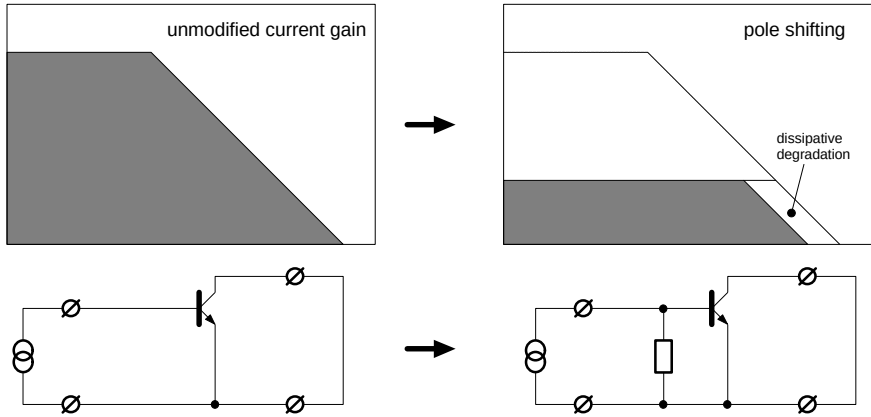


Fig 5.5 Example of an amplifier stage with intertwined first order profiled compensation. The two asymptotic Bode plots of the current gain illustrate symbolically that the compensated transfer originates from a virtual cascade of the uncompensated transfer and the desired compensation function.

Intertwined compensation by pole splitting

Figure 5.6 shows an example of an intertwined second order profiled compensator. The additional resistor of figure 5.5 is replaced by a series RC-network for maintaining the current gain at low frequencies. The asymptotic Bode plots illustrate that this intervention replaces the uncompensated transfer with a mid-band pole by a compensated transfer with two poles and a zero. The poles are positioned at either sides of the pass-band and the zero somewhere between these two. This is called *pole splitting*.

The preservation of low frequency gain is one of the benefits of pole splitting, compared to pole shifting. Another benefit is that the position of the generated zero is a free choice. This reduces the number of compensation zeros that is deemed to take the place of a pole, which is a typical restriction of intertwined solutions. The location of no more than one compensation zero is committed to the position of an uncompensated pole.

The second example in figure 5.6 performs an equivalent transfer function with local feedback. The benefit of this approach is that the preservation of the high frequency current gain may be achieved without inductor. A small capacitor, shunted with resistor R_z , relaxes the local feedback for high frequencies, and in this way relaxes the reduction of high frequency gain. The use of a parasitic transistor capacitance, and a properly matched local feedback network may fulfill all requirements.

The launch of similar local feedback loops with both inductors and capacitors facilitates the construction of higher order compensators with complex poles and zeros.

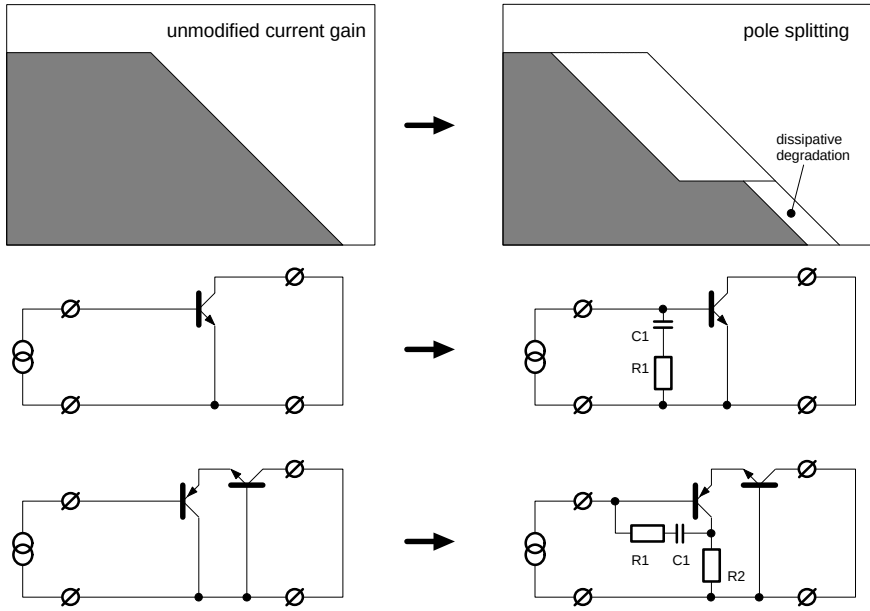


Fig 5.6 Example of amplifier stages with intertwined second order profiled compensation. The two asymptotic Bode plots of the current gain illustrate symbolically that the compensated transfer originates from a virtual cascade of the uncompensated transfer and the desired compensation function.

5.2.3. Phantom compensation techniques in the feedback network

The feedback factor is another parameter in the superposition model that is accessible for effective compensation. This type of compensation is identified as *phantom* compensation.

The name *phantom* refers to the property that the insertion of new phantom zeros do not affect the position of the zeros in the effective-aperture and do not expand the number of zeros. This remarkable feature facilitates a way to manipulate effective-aperture with the addition of phantom zeros to the loop but without the addition of new singularities to the effective-aperture.

To make this statement stick, the effective-aperture is written out in polynomial form:

$$\hat{H}(s) = H_0(s) \cdot H_{cs}(s) \cdot H_{cp}(s) = \frac{T_0(s)}{N_0(s)} \cdot \frac{T_{cs}(s)}{N_{cs}(s)} \cdot \frac{T_{cp}(s)}{N_{cp}(s)} = \text{loop gain}$$

$$H_{sa}(s) = \frac{1}{H_{cp}} \cdot \left(\frac{\hat{H}}{\hat{H}-1} \right) = \frac{T_0 \cdot T_{cs} \cdot N_{cp}}{T_0 \cdot T_{cs} \cdot T_{cp} - N_0 \cdot N_{cs} \cdot N_{cp}} = \text{effective-aperture}$$

Phantom zeros, which are the roots of the polynomial $T_{cp}(s)$, contribute to the denominator of the effective-aperture however are conspicuous by absence in the numerator. Note that this does not hold for the poles of a phantom compensator.

Phantom compensation is basically the addition of zeros in the loop to affect the undesired poles in the passband. The exact location of these zeros is discussed in section 5.3, however in general they are located at the high end of the passband. This subsection restricts itself to the overall principle.

Unlike profiled zeros, the addition of (passive) phantom zeros is not necessarily accompanied with attenuation of the loop gain. This is because the feedback network is commonly an attenuator itself, and the addition of zeros is nothing more than relaxing this attenuation.

Figure 5.7 shows the impact of this statement. The first example shows a feedback factor of $\beta=1/25$, cascaded with a first order (passive) phantom compensator. The second example shows an equal feedback factor, but cascaded with two (complex) phantom zeros. The phantom zeros causes the magnitude response to increase with frequency until all feedback is maximized. As a result, the phantom zeros come in pairs with parasitic poles, and these poles do not improve the overall transfer. The more phantom zeros are inserted, the sooner this limit is reached and the lower the corner frequencies will be of the parasitic poles.

The most conspicuous aspects of phantom compensation are the lack of any loop gain reduction and the ability to keep all parasitic poles out of the passband. Because of these features, phantom compensation is preferred to profiled compensation.

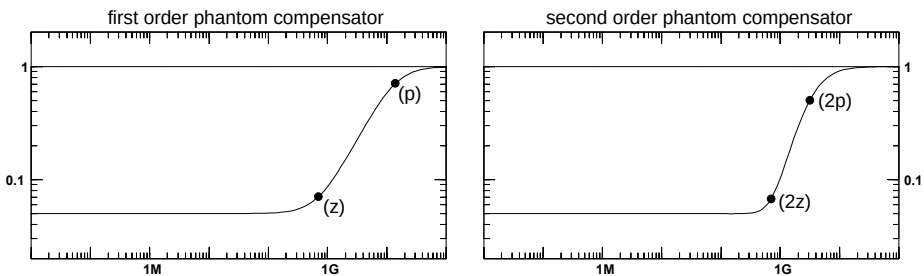


Fig 5.7 Typical example of the magnitude response of a feedback network with (phantom) compensation. Lowband and midband frequencies are not affected, preserving the intended feedback factor. Since the feedback network is basically an attenuator, the compensator reduces the attenuation for highband frequencies to create the compensation zeros. Note the parasitic pole(s) that occur when all attenuation is neutralized.

Figure 5.8 shows various examples with phantom compensation in optical receivers. They are all terminated with a capacitive load, because a first order approximation of a practical load is more often capacitive than inductive.

The shunt capacitor, added to the transimpedance receiver in figure 5.8a, increases the RF-signal flow from output to input. This increases the feedback factor and yields a phantom zero.

The series resistor, added to example 5.8b, inhibits the RF-signal flow from the output of the amplifier to the load impedance, which is another way to increase the feedback. The combination of both the serial and the shunt resistor generates two phantom zeros.

In very wideband receivers, the parasitic shunt capacitor across the transimpedance feedback resistor causes a feedback factor with an annoying corner frequency inside the passband. The feedback network in figure 5.8c equalizes the feedback factor by a counter-productive RC-network. The serial resistor, that is added to this example, obstructs this equalization, and facilitates a phantom zero.

The last example in figure 5.8d is an optical receiver with current-current feedback, in which the current flow through the two floating output terminals is fed to the input by a capacitive current divider. This current division is frequency independent. The serial resistor in the grounded portion of this divider causes an unbalanced current division for high-band frequencies and yields a phantom zero. An additional inductor causes two complex phantom zeros.

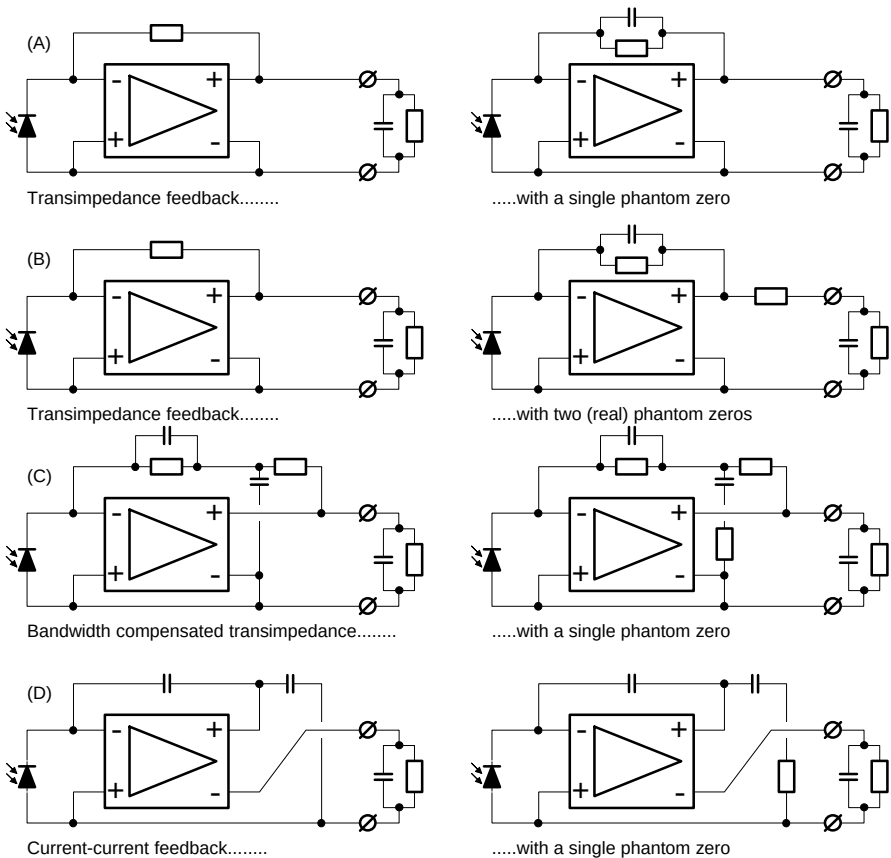


Fig 5.8 Examples of phantom compensation in various types of optical receivers. All uncompensated examples are drawn on the left side and the associated compensated examples on the right.

All examples of figure 5.8 suffer from parasitic poles in the transfer of the phantom compensation. The capacitive load and the parasitic feedback capacitors prevent the phantom compensator from being optimal, and introduce an additional penalty. This

penalty is configuration dependent and limits the launch of phantom compensation in a particular situation.

5.2.4. Conclusions

In conclusion, some restrictions of optimizing circuit simulators are discussed, when used for compensating feedback amplifiers. A formal compensation synthesis is proposed, that is isolated from how the compensation should be implemented. It enables future circuit simulators to inform designers what compensation transfer is required for proper compensation.

The properties of passive compensation networks in the forward amplifier (profiled compensators) are analyzed. The transfer of optimal (passive) profiled compensators is characterized by a 'dip' at midband frequencies and unity gain at low and high frequencies. The exact transfer is discussed in section 5.3.

Profiled compensators are usually intertwined with the remaining circuitry. These solutions, however, have fewer degrees of freedom than distinct solutions. Some intertwined solutions are summarized using well-known pole-shifting and pole-splitting techniques.

The properties of passive compensation networks in the feedback network (phantom compensators) are analyzed. The transfer of an optimal (passive) phantom compensator is characterized by unity gain at low and midband frequencies, and high gain at high frequencies. As a result, adequate passive solutions are always intertwined with the feedback circuitry.

5.3. Compensation synthesis for feedback amplifiers

The previous section analyzed the transfer of optimal profiled and phantom compensators. This section 5.3 reverses this process, and synthesizes the compensation transfer required to achieve a effective-aperture with an optimal transfer.

The synthesis of the compensation, starting from a specified effective-aperture and available loop gain, is a complicated process. Innumerable solutions do exist but they are very difficult to find on a trial and error base.

Nordholt [406] proposed a structured method to design the compensation for the special case that the loop gain is an all-pole transfer function with two or three poles. Stoffels [510] automated this design method, to generate a circuit diagram and its compensation, guided by user specifications. Essential for the above methods is that the annoying effects of parasitic poles and zeros is adequately reduced by additional sub-circuits (buffering, cascoding).

At low frequencies, e.g. below $f_T/10$ for second order loops, this approach is successful. A fully automated synthesis of amplifiers has been demonstrated [510] for frequencies up to 1 MHz (second order loop). However, these methods may fail for increasing bandwidth demands, e.g. above 200 MHz.

Section 5.2 discussed the use of circuit optimizers, to extend the applicability of these structured methods. However, a serious problem arises when optimizers are as well ineffective in achieving the desired frequency response.

These wideband feedback amplifiers have in common that they suffer from loops with dominant zeros, from RHP-zeros and from many parasitic poles and zeros that cannot be ignored. The order of the loop transfer function is then too high for compensation of *all* relevant singularities.

No solutions were found in the literature on controlling these loops.

This section 5.3 introduces a powerful algorithm for compensation synthesis that has been developed during this study. It provides the transfer function that is required for additional compensation.

The proposed method can handle loops with dominant RHP-zeros and loops that suffer from parasitic singularities. This algorithm predicts the required compensation, in terms of poles and zeros, where conventional (manual) methods fail.

Finally, examples are presented in this section that illustrate the power of the synthesis algorithm, and may serve as a template for manual synthesis.

5.3.1. Bandpass synthesis

Consider a feedback amplifier topology, without stability compensation, for instance designed using Nordholt's method [406]. A first requisite for automated compensation synthesis is a mathematical description of the desired effective-aperture passband. This mathematical description requires the specification of the *generic passband* and the evaluation of the *specific passband*. Note that an unambiguous passband specification is not equivalent to a unique compensation solution.

generic passband specification

Compensation synthesis starts from the specification of a generic passband. The generic passband, is a normalized transfer function that serves as a scale model for the effective-aperture. Its dc-gain and its asymptotic corner frequency are one by definition. Further, the generic passband is defined to be a minimum-phase transfer function, which means that all singularities are positioned in the left complex half-plane.

A practical effective-aperture is a low-pass transfer function that must be smoothed and optimally flattened. Non-coincident pole-zero pairs are undesirable in the generic passband because they introduce additional ripple. The generic passband is therefore primarily an all-pole transfer function, whose order is chosen as low as possible.

Select a suitable all-pole transfer function, for instance a Bessel, a Butterworth or a Chebyshev transfer function, based on the bandpass analysis presented in section 5.1.3. The pole-zero patterns in figures 5.9a and 5.3 are typical for these transfer functions.

specific passband evaluation

The specific passband serves as a properly dimensioned model for the effective-aperture. The specific passband is evolved from the generic passband through appropriate magnitude, bandwidth and delay. The scaling operations facilitate matching the effective-aperture to the design goal, with respect to the compensation limits of the loop.

The specification must take account of the limitations of passive compensation, so that the specified bandwidth should match the available loop bandwidth. We identify this as magnitude and bandwidth scaling. One way to modify the feedback loop to match this passband specification is removing the undesirable poles by introducing coincident compensation zeros.

Although the opposite applies for LHP-zeros, RHP-zeros cannot be compensated using RHP-poles due to stability considerations. This study has resulted in novel techniques on handling RHP-zeros by treating them as *pseudo delay*. Building on this concept, a method has been developed in this study that adds an appropriate delay to the passband specification to satisfy the compensation limitations of the loop. We identify this as delay scaling.

Let us recall what modifications to the specific passband are possible that preserve the magnitude and bandwidth scaling. A typical passband specification is illustrated in figure 5.9.

- Figure 5.9a is indicative for the initial pole-zero pattern of the generic passband, resulting from appropriate scaling in magnitude and bandwidth.
- The addition of coincident pole-zero pairs, as in figure 5.9b, will not affect the transfer properties. Therefore, they are allowed in a passband specification when required.
- Addition of mirrored pole-zero pairs, such as in figure 5.9c, contributes a frequency-dependent phase shift. This is equivalent to pseudo delay within a limited frequency band. These pole-zero pairs are allowed in a passband specification when unavoidable and when their magnitude is equal or higher than the width of the passband.

All other pole-zero patterns are in principle undesirable.

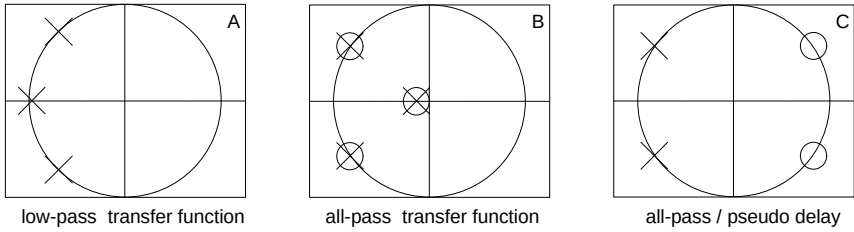


Fig 5.9 Example of the basic pole-zero patterns that composes the specific passband of the effective-aperture. It comprises of: (a) a low-pass all-pole transfer function, (b) an all-pass transfer function with coincident pole-zero pairs and (c) an all-pass phase-shifting transfer function with mirrored pole-zero pairs.

What pole-zero pattern is most appropriate for meeting the passband specification? From the definition of the effective-aperture it is observed that the zeros in the profiled loop gain show up at the same location as in the effective-aperture. Since the specific passband is preferably an all-pole transfer function, these zeros call for cancellation by coincident poles. They should not pop up in the passband specification, or at most in the shape of a coincident pole-zero pair, as drawn in figure 5.9b.

Unfortunately, this is not feasible for RHP-zeros. We assume that the best approach for dealing with RHP-zeros is to let them function as pseudo time delay. They should therefore appear in the passband specification in combination with poles that are mirrored with respect to the imaginary axis, as illustrated in figure 5.9c.

In short, all zeros in a compensated effective-aperture must appear in the passband specification, in combination with poles at their *equivalent minimum phase* position. These poles are defined as poles that create coinciding pairs with LHP-zeros and create mirrored pairs with RHP-zeros. The mirrored pole-zero pairs contribute a constant phase delay to the effective-aperture over a wide frequency interval. The coinciding poles and zeros neutralize each other.

From the definition of the effective-aperture it is recognized that the poles in the loop gain show up at a different location as in the effective-aperture. They represent the remaining restrictions on the feedback loop, and there is nothing left than to accept these poles in the specification. The best that can be achieved is to position them as in the generic passband, properly scaled in bandwidth.

All scaling steps together result in the following transformation of generic passband into specific passband:

- *Magnitude scaling* contributes an exact match with the effective-aperture at zero frequency. This scaling factor can be evaluated simply. It equals $(g)/(g-1)$, in which g is the dc-gain of the loop.
- *Bandwidth scaling* contributes a close magnitude match to the effective-aperture for all frequencies within the passband. It requires the calculation of the available bandwidth (ω_b), starting from a dominant magnitude match for infinite frequencies. This calculation has been described in section 5.1.2.
- *Delay scaling* is the finishing touch and contributes a full match of both magnitude and phase. It fits the minimum-phase transfer of the generic passband, and the non-minimum-phase transfer of a effective-aperture with RHP-zeros. When $T_0(s)$

represents the polynomial with all profiled zeros of the loop gain and $T_0(s)$ represents the associated polynomial with all roots at their equivalent minimum phase position, than the transfer function $T_0(s)/T_0(s)$ provides the (pseudo) delay requested.

The final specification of the specific passband $A(s)$ results in the following mathematical form:

$$A(s) = \frac{g}{g-1} \cdot \frac{T_0(s)}{T_0(s)} \cdot \frac{1}{B_0(s/\omega_0)} = \text{specific passband}$$

In this expression, the polynomial $B_0(s)$ refers to the numerator of the chosen generic passband. The quantity g refers to the dc-loop-gain and performs the magnitude scaling. The quantity ω_0 refers to the available bandwidth and performs the bandwidth scaling. The polynomial $T_0(s)$ refers to the denominator of the profiled loop gain, and its roots produce the zeros of the effective-aperture. The polynomial $T_0(s)$ refers to the minimum phase equivalent of $T_0(s)$, and its roots produce the poles in the effective-aperture that accompanies the zeros. The rational function $T_0(s)/T_0(s)$ forms an all-pass transfer function with a pseudo delay component.

5.3.2. Compensation algorithm

A unique solution for the compensation of a feedback loop does not exist. Nevertheless, an algorithm should lead to a unique result. In this section, the prior conditions are established that will yield the desired unique solution, which may be linked to practical compensation considerations.

Next, a powerful compensation algorithm is introduced, in two steps. The *full* compensation algorithm holds for the special case that all poles and zeros in the loop are compensatable. The *dominant* compensation algorithm generalizes the synthesis for loops with parasitic poles and zeros.

prior conditions for a unique compensation solution

Consider the synthesis of a particular loop gain requirement, starting with an n^{th} -order specific passband transfer function. The synthesis procedure should find an expression for the loop gain matching with $(n+1)$ constants, for instance organized as the poles and the dc-gain of the predefined specific passband.

Reliable loop gain synthesis starts from the assumption that the desired specific passband is compatible with the dc-gain and the available bandwidth of the loop. The reduced synthesis should match with the $(n-1)$ remaining constants, for instance organized as the parameters of the generic passband transfer function. In practical terms: the compensation of an n^{th} -order loop requires at least $(n-1)$ compensation interventions. The loop gain has at least n poles in the loop, and that is one more than is required for a unique solution. The degrees of freedom increase in number when (phantom) zeros are permitted in the solution for a compensated loop gain. On one hand, it is not difficult to find just some solution, however a practical implementation of this solution will not fulfill the requirements of practical feedback loops. On the other hand, it is very difficult

to find the remaining unique solution using manual synthesis, imposed by prior conditions.

The synthesis algorithm proposed here exactly solves $(n-1)$ loop parameters in the case that all phantom zeros (m) in the feedback network and some poles $(n-m-1)$ in the profiled loop gain are variable. This number is the exact number that is required to yield a unique compensation solution.

All remaining loop parameters are arbitrarily predefined, which is equivalent to fixing all zeros and all remaining poles in the uncompensated loop gain. These prior conditions link up nicely with practical compensation problems because most singularities in the loop maintain their position after a successful sequence of compensation interventions.

full compensation

Initially, the question which loop gain $\hat{H}(s)$ yields a effective-aperture with predefined generic passband is considered in the case that dc-gain of the loop and some poles and zeros are predefined? To find this solution, the compensated and uncompensated loop gain are described in the following rational form:

$$H_0(s) = g \cdot \frac{T_0(s)}{N_{00}(s) \cdot N_{0x}(s)} = g \cdot \frac{T_0(s)}{N_0(s)} = \text{uncompensated loop gain}$$

$$\hat{H}(s) = g \cdot \frac{T_0(s) \cdot T_x(s)}{N_{00}(s) \cdot N_x(s)} = g \cdot \frac{\hat{T}(s)}{\hat{N}(s)} = \text{compensated loop gain}$$

The polynomials $N_0(s)$ and $T_0(s)$ represent the uncompensated poles and zeros and the constant g represents the dc-gain of the loop. All zeros (T_0) and some poles (N_{00}) are predefined, which means that the compensation may not modify the position of these singularities. The remaining poles (N_{0x}) are selected for modification by the compensation procedure.

The polynomials $\hat{N}(s)$ and $\hat{T}(s)$ represent the compensated poles and zeros, of which $N_{00}(s)$ and $T_0(s)$ are the predefined singularities, and $N_x(s)$ and $T_x(s)$ are the unknown singularities that will be extracted by the synthesis process. *All* roots of $N_x(s)$ are originated in the forward amplifier (profiled poles) and *all* roots of $T_x(s)$ are compensation zeros in the feedback network (phantom zeros). All polynomials are normalized in magnitude to ensure that $N_{00}(0)=T_0(0)=1$ and $N_x(0)=T_x(0)=1$.

Appendix J describes the mathematics of the algorithm proposed for the extraction of the polynomials $N_x(s)$ and $T_x(s)$. It is based on the weighted polynomial division of appendix G, that is an integral part of this study. A symbolic representation for this process, annotated as a function of four input variables and two output variables, yields:

$$[N_x(s), T_x(s)] = \text{CompSynthesis}(g, N_{00}(s), T_0(s), B_0(s/\omega_0))$$

It requires the generic passband polynomial $B_0(s)$, scaled in bandwidth by the available bandwidth ω_0 of the uncompensated loop gain. The order (m) of $N_{00}(s)$ is the number of profiled poles (m) that must maintain their positions, and this presets the required number of phantom zeros (m) that will be produced by the output polynomial $T_x(s)$. The order (n) of $B_0(s)$ fixes the number of profiled poles $(n-1-m)$ that must be modified and

that will be produced by the output polynomial $N_x(s)$. The other two input parameters specify the predefined poles and zeros and the loop gain.

The two extracted polynomials, $N_x(s)$ and $T_x(s)$, produce the transfer functions of the compensating networks. From the definition of compensated and uncompensated loop gain, we obtain:

$H_{cp}(s)$	$= \frac{T_x(s)}{1}$	$= \textit{phantom compensator, with order } (m)$
$H_{cs}(s)$	$= \frac{N_{0x}(s)}{N_x(s)}$	$= \textit{profiled compensator, with order } (n-m-1)$

The zeros in the profiled compensator cancels all the selected poles (N_{0x}) of the loop. This is because $N_{0x}(s)$ originates from the uncompensated loop. In practice, this is equivalent to the replacement of some poles in the loop, by some new, precisely positioned, compensation poles in $N_x(s)$.

Furthermore, the algorithm ensures that some of the remaining poles in the profiled compensator are used to cover all LHP-zeros of the uncompensated loop. This minimizes the effective order of the compensated loop gain.

An inaccurate evaluation of the available bandwidth causes a difference between the highest coefficients of the polynomials $N_{0x}(s)$ and $N_x(s)$. This is because $N_{0x}(s)$ is not an input parameter for the algorithm. As a result, the better that ω_0 is estimated, the better that $H_{cs}(s)$ approximates one for infinite frequencies.

dominant compensation

Up to now, it has been understood that all the loop singularities are compensatable. This implies that all singularities are within or close to the bandwidth circle, and that the compensation solution will be exact. Restrictions to the zeros of the intertwined compensation solutions are no longer of further concern with full compensation, since all profiled zeros coincide with uncompensated poles in the loop.

Full compensation is feasible when parasitic singularities are irrelevant, which is typical of low frequency design. In the case of significant parasitic singularities, a dominant approximation of the loop is required to keep control over the transfer order of the compensation solution. This requires the replacement of all singularities by a smaller set with (other) singularities that approximate the loop within a predefined frequency band. Thereafter, apply the overall compensation synthesis to the dominant singularities and proceed as described above.

The necessity to correct the phase of the loop gain is concentrated near the maximum available bandwidth ω_0 . This implies that the accuracy of the approximation must be optimized near ω_0 . Start therefore from the estimation of the available bandwidth ω_0 and perform the dominant approximation from dc to ω_0 . Especially the refinement step of the accuracy near ω_0 , as described in section 4.5.3, is important because accuracy failures will significantly deteriorate the overall synthesis result.

The introduction of phantom zeros by the synthesis algorithm results in additional parasitic poles in the loop. Their location is application dependent and is discussed in section 5.2.3. Since they were ignored in the synthesis, an iterative approach based on our synthesis procedure is required:

- Initialize the compensation synthesis based on an uncompensated loop.
 - Estimate the additional parasitic poles associated with phantom compensation.
 - Add these poles to the uncompensated loop and recalculate the compensation.
- Repeat these iterative steps as often as is required to stabilize the predicted results.

A practical implementation of the compensation prefers the use of intertwined compensation networks. Unfortunately, the compensation zeros required do not coincide with the (primary) poles in the loop, when applying dominant compensation. Instead, they coincide with those dominant poles that differ in position from the primary poles. As a result, the use of intertwined compensation schemes may restrict or in fact undo the compensation performance. This is because the intertwined compensation solutions are directly associated with compensation zeros coinciding with primary poles.

A practical solution to this problem is a complementary mathematical adjustment of the predicted compensation transfer function. This adjustment step must modify all compensation singularities to force one or more profiled compensation zeros to a predefined position and to conserve a fair approximation of the required transfer function.

This is illustrated in the following example of a third order profiled compensator. The right side of the equation is the original result that is predicted by a dominant compensation synthesis and the left side is the modified result

$$\frac{(1-s/z_{10}) \cdot (1-s/z'_2) \cdot (1-s/z'_3)}{(1-s/p'_1) \cdot (1-s/p'_2) \cdot (1-s/p'_3)} \approx \frac{(1-s/z_1) \cdot (1-s/z_2) \cdot (1-s/z_3)}{(1-s/p_1) \cdot (1-s/p_2) \cdot (1-s/p_3)}$$

In this example, all singularities are modified to change the zero z_1 into z_{10} and to facilitate a fair agreement between the left hand expression and the right hand expression. This process is equivalent to:

$$\frac{(1-s/z'_2) \cdot (1-s/z'_3)}{(1-s/p'_1) \cdot (1-s/p'_2) \cdot (1-s/p'_3)} \approx \frac{(1-s/z_1)}{(1-s/z_{10})} \cdot \frac{(1-s/z_2) \cdot (1-s/z_3)}{(1-s/p_1) \cdot (1-s/p_2) \cdot (1-s/p_3)}$$

Of course there exist no exact solution and this clarifies the statement that an intertwined compensation may restrict the overall compensation performance. To find a fair approximation, an appropriate deflation technique can be used, such as the pole-zero cancellation algorithm of section 4.5.2. In an emergency, the curve fit algorithm for rational functions can give a solution.

Note that pole splitting techniques are preferred to pole shifting techniques, because they entail more degrees of freedom. Pole splitting techniques make (at least) one additional zero available for compensation, which is not restricted to the location of another (primary) pole in the loop.

5.3.3. Examples of loops with full compensation

Practical feedback design cannot rely on the blind use of automated synthesis tools. Compensation synthesis always requires some manual assistance, and root-locus plots are quite suitable for the assessment of the synthesized result. These plots give information on the loop gain, the effective-aperture, and the manner in which these two are interrelated.

A root-locus plot is a composite description of an infinite number of pole-zero patterns, with a minor difference between them. It is commonly practice to vary the dc-gain parameter from zero to infinity although there is no good criterion for selecting this parameter for the *best* available choice. These plots are useful because they are amenable to *pattern recognition* techniques. The effectiveness of this manual technique is highly dependent on the correlation of interpretative graphic rules for associative root-locus behavior with distinct changes in the loop.

This subsection summarizes various typical examples of root-locus plots to acquaint the reader with a number of commonly known graphic rules based on pattern recognition.

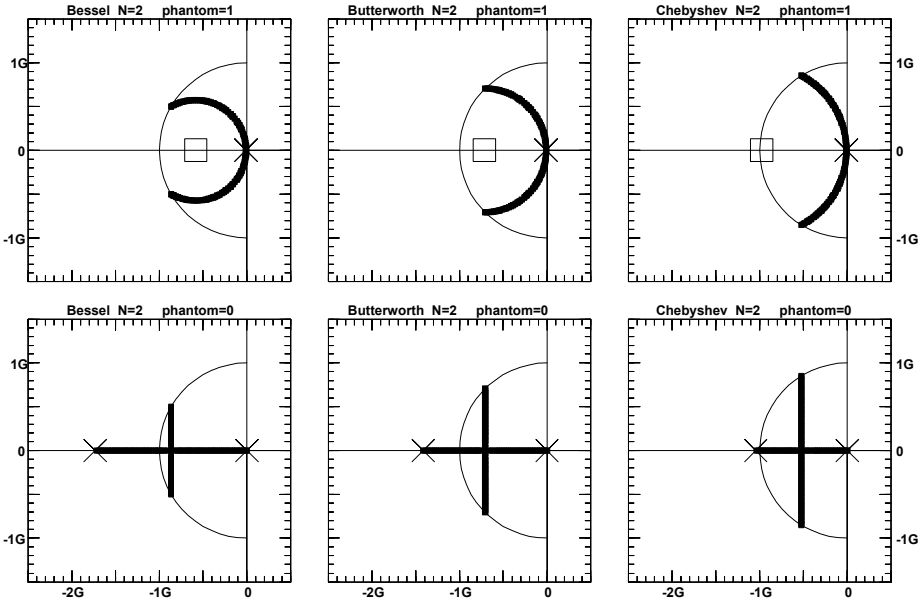


Fig 5.11 Examples of finite root-locus plots for compensated second order feedback loops with all dominant poles. In these examples, the poles near the origin are at 10 MHz and the feedback bandwidth is 1 GHz.

Figure 5.10 to 5.12 show various *finite* root-locus plots for a different number of phantom zeros and for various generic passband specifications. Finite root-locus plots are slightly different from the usual infinite root-locus plot. Finite root-locus plots illustrate the positions of all poles and zeros of the virtual-aperture, while the dc-gain is varied continuously from its *current* value down to zero.

The points of the curved lines originate from the poles of the following transfer function, while α varies from zero to one:

$$A'_{la}(s) = \left(\frac{\alpha \cdot \ddot{H}}{\alpha \cdot \ddot{H} - 1} \right) \quad \text{with} \quad \{ \alpha \in \mathbb{R} \mid 0 < \alpha \leq 1 \}$$

All infinite numbers of poles are marked with dots, which yields the curved lines. The 'beginning' of the lines are marked with cross markers that represent the poles of the loop gain in a usual way. The 'end' of the lines mark the positions of the poles of the

virtual-aperture, which is equivalent to the positions of the poles of the effective-aperture. The phantom zeros are represented by square markers, while all remaining zeros (if any) are represented with circular markers.

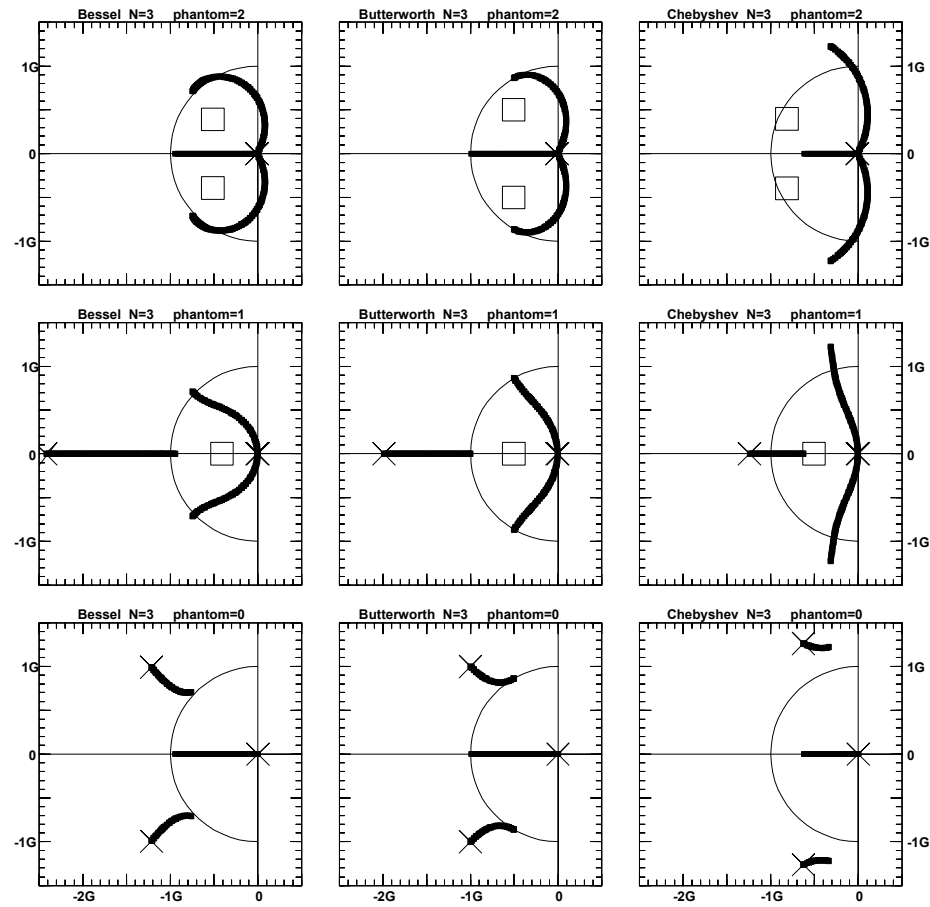


Fig 5.11 Examples of finite root-locus plots for compensated third order feedback loops with all dominant poles. In these examples, the poles near the origin are at 10 MHz and the feedback bandwidth is 1 GHz.

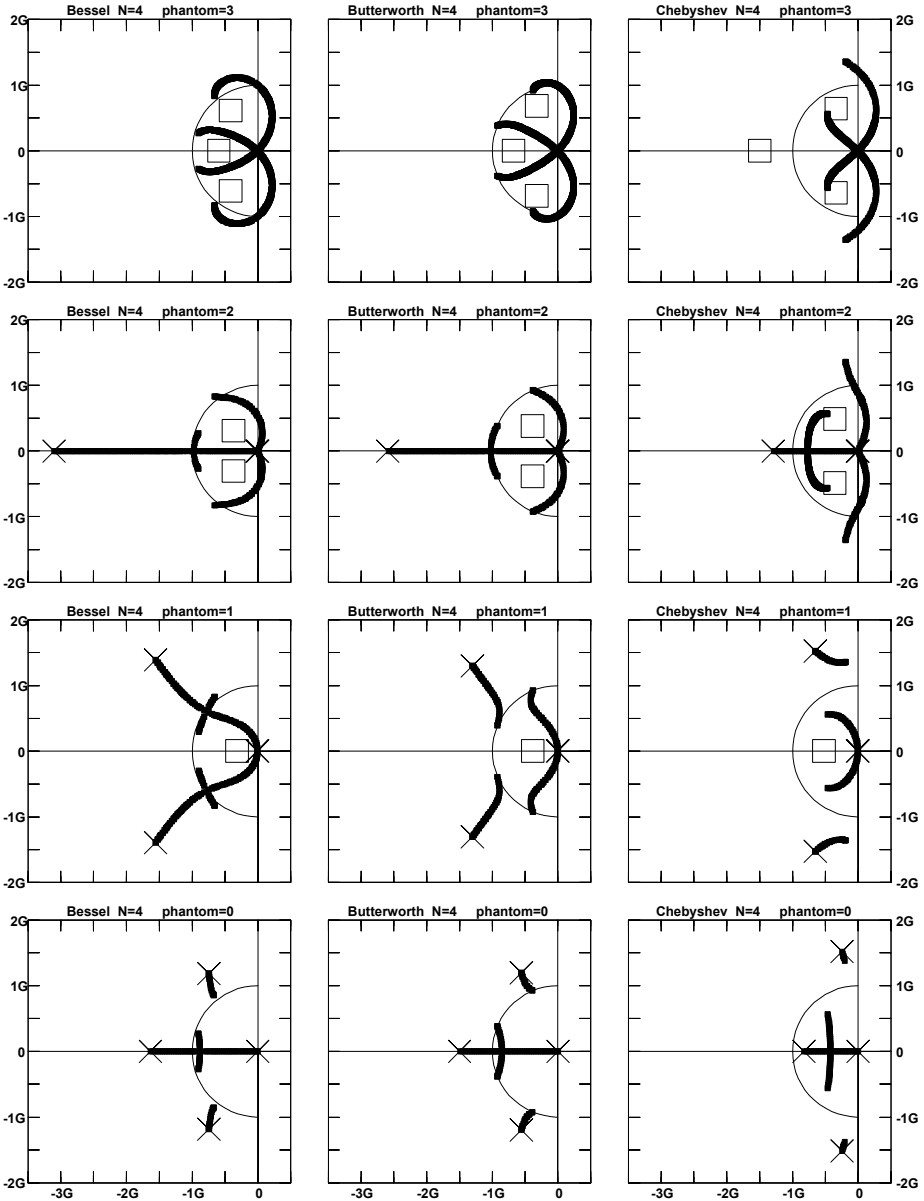


Fig 5.12 Examples of finite root-locus plots for compensated fourth order feedback loops with all dominant poles. In these examples, the poles near the origin are at 10 MHz and the feedback bandwidth is 1 GHz.

The pole-zero patterns in figure 5.10 to 5.12 are typical of compensated loops, in which all amplifier stages provide a dominant pole of which the corner frequency is significant smaller than the available bandwidth. These are all idealized situations; an example from practice would never look that perfect.

When plotted linearly, the uncompensated pole positions are close to the origin of the pole-zero pattern, with respect to the radius of the bandwidth circle. In all these examples the corner frequencies of the uncompensated poles are fixed at 1% of the available bandwidth. The exact position is not relevant because poles located at 10% or 0.1% will result in nearly similar root-loci. The remaining pole positions are adjusted by the compensation, when required.

Various rules of thumb may be extracted from these typical examples. At least the commonly known rules are of value, such as:

- All root loci start from the positions of the poles in the loop.
- All extrapolated root loci end at the positions of the zero in the loop or end in infinity, when the loop gain is expanded to infinity.
- The poles and zeros pass for virtual magnets that 'repel' the root locus lines with poles and 'attract' the root locus lines with zeros.

Further, these plots show that:

- All compensating poles and all phantom zeros are located *near* the bandwidth circle, when a *logarithmic* scale was used.
- Most phantom zeros are located *inside* the bandwidth circle.
- Most compensating poles are located *outside* the bandwidth circle.
- A magnitude response that emphasizes the high end of the passband (Chebyshev low-pass transfer) requires an increase in magnitude for the phantom zeros or a decrease in magnitude for the compensating poles. The converse holds for a passband with a smoothed magnitude response (Bessel).
- Compensation using the insertion of *two* or more phantom zeros requires a complex pair of zeros.
- Compensation by displacement of *two* or more poles requires a complex pair of poles.
- The insertion of *two* or more phantom zeros requires a complex pair of zeros.

These empirical rules are valuable for a first guess on the location where the phantom zeros should be inserted or whereto the poles should be moved.

5.3.4. Examples of loops with dominant compensation

The typical examples in section 5.3.3 were intended to get experience in the interpretation of computer generated root-locus patterns. They are applicable for manual compensation synthesis in a frequency band in which primary poles and dominant poles are almost equal.

Manual synthesis of complex feedback loops, such as the examples in section 4.4.3, is next to impossible, and must rely on automated synthesis. This subsection 5.3.4 demonstrates the compensation of realistic feedback loops.

We start from an arbitrary example of a loop covering three cascaded amplifier stages. In section 4.4.3 three examples of transimpedance feedback amplifiers are discussed to show their associated loop gain as Bode plot and as pole-zero pattern. The loop of example *b* (cascoded amplifier stages) was modeled with six poles and three zeros up to 3 GHz.

At first, the available bandwidth was estimated and rated at 490 MHz. Next, the loop was deflated with a dominant approximation of four dominant poles, and one dominant zero. Because this dominant zero is located in the *right* half plane, further deflation by pole-zero cancellation is out of the question. Finally, the synthesis algorithm is applied to the deflated loop, to produce one of the results of figure 5.13.

Three different compensation solutions were evaluated for a Butterworth passband:

- No phantom zeros and a 4th order profiled compensator
- One phantom zero and a 3rd order profiled compensator
- Two phantom zeros and a 2nd order profiled compensator

The plots in figure 5.13 show that the more phantom zeros are available for use, the less attenuation is required in the profiled compensator, and the lower its transfer order will be. For convenience, the parasitic poles are ignored that are generated in the phantom compensator.

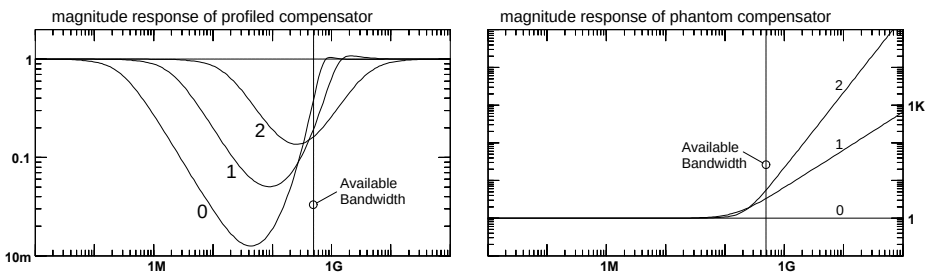


Fig 5.13 Examples of the magnitude response of the required compensation for a dominant 4th order feedback loop. The numbers 0,1,2 refer to the number of phantom zeros that are inserted in the loop.

The effective-aperture that arises from the compensated loop is shown in figure 5.14. In this plot, the dominant compensation is combined with all the poles and zeros of the uncompensated loop. As a result, the magnitude responses differ from the specified passband, especially *above* the passband.

Note that the ripple on the deviation from the specified passband increases with the number of phantom zeros. This illustrates that phantom compensation is not *always* preferable to other compensation techniques, as was stated by Nordholt [406:p175]. We

emphasize that the efficiency of phantom compensation is deteriorated by its own parasitic poles.

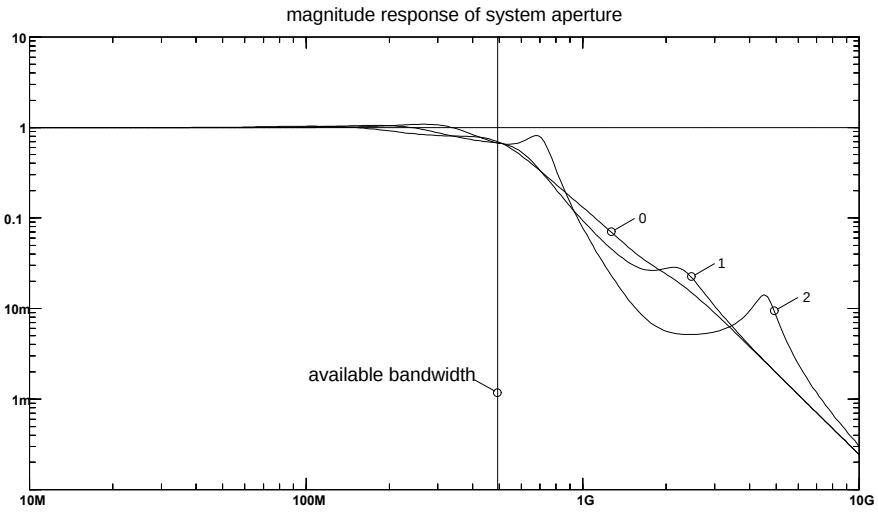


Fig 5.14 Examples of the effective-aperture that is associated with the compensation of figure 5.13. The numbers 0,1,2 refer to the number of phantom zeros that are inserted in the loop.

All the feedback loops mentioned here are stable. This is demonstrated in the root-locus plot of figure 5.15. Both uncompensated and compensated loops are plotted, for the situation that one phantom zero is inserted.

The mirrored positions of some poles, with respect to the RHP-zeros, are all out of recognition. This is because those positions would have captured in a modified root-locus plot where the singularities are replaced by their dominant counterparts. The complexity of the current root-locus plot illustrates why manual synthesis will fail for these loops.

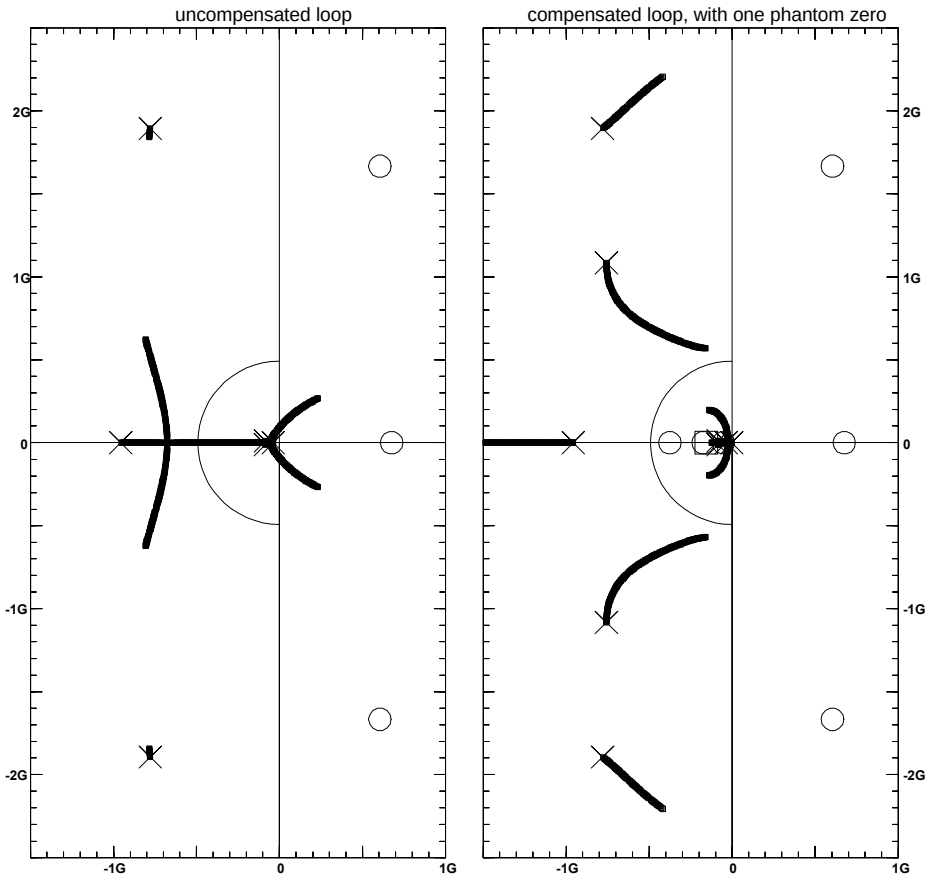


Fig 5.15 Example of a reliable root-locus plot. The uncompensated loop was modeled with six poles and three zeros. The compensation has added one phantom zero and a third order profiled compensator.

5.3.5. Conclusions

In conclusion, the difficulties are discussed that arise when the order of the loop transfer function is too high for compensation of *all* relevant singularities. Conventional compensation methods will then be inadequate. The assistance of circuit optimizers may then become crucial, however they require realistic design goals.

A bandpass synthesis algorithm is developed that provides an accurate estimation of the effective-aperture, when optimal passive compensation is applied. This transfer function, the *specific passband*, is extracted from the poles and zeros in the loop using the aperture analyses of section 5.1 and the deflation algorithms of section 4.5.

The bandpass synthesis is robust and is proofed against RHP-zeros in (right half plane). The algorithm let these zeros show up as pseudo delay. This means that the pole-zero pattern of the predicted effective-aperture includes a mirrored pole-zero pair.

The synthesized specific passband is of direct use as input for automated circuit optimization.

A compensation synthesis algorithm has been developed, in terms of poles and zeros. As input, it requires the result of the specific passband calculation and the number of phantom zeros that is required. As output, it provides an accurate estimation of the required transfer functions of the two compensation networks: profiled compensator (in forward amplifier) and phantom compensator (in feedback network).

The designer can use this for assessing the feasibility of the compensation. If it is feasible then it is of direct use in designing the compensation network configuration. This is significant simpler than designing it from scratch.

An optimizer may fine tune the compensating elements for optimal performance.

The performance is demonstrated of the compensation synthesis algorithm, in the exceptional case that *all* poles of the loop are compensated (full compensation). These examples are primarily intended for *pattern recognition* when assessing root-locus patterns.

Furthermore, the performance of the compensation synthesis algorithm is demonstrated in the realistic situation that the loop has many poles and zeros, of which a few of them are dominant (dominant compensation). It demonstrated that the loop of example "b" in section 4.4.3 is compensatable, although higher order compensation networks are required.

Chapter 6

Feedback design example: lightwave receivers

The previous chapters have discussed various theories and tools (measurement methods, algorithms) to analyze the response of circuits with feedback loops. They discussed the synthesis of compensation networks for stable feedback loops. When designing circuits, structured design methods are to be preferred, as in example the feedback network synthesis method proposed by Nordholt [406]. Parasitic effects deteriorate the overall performance, and they may even endanger the loop stability in the case of wideband feedback circuits.

When using *estimated* device models, it is utopian to believe that higher order wideband feedback loops designed from scratch will perform as expected from the initial analysis. In practice, *iterative* synthesis¹ must be used to improve the initial design. Various relevant parameters need to be extracted from the measured data (or *reliable* simulation) to gain insight into annoying parasitic effects. Study of these extracted values leads to modifications for improving the overall design. Where required, these steps must be repeated, converging eventually to optimal performance.

This chapter discusses the applications of our design tools using an example chosen from practice. It illustrates the suitability of our tools to the support of iterative synthesis of lightwave receivers, designed for more than 1 GHz in bandwidth .

An introductory section 6.1 reviews various receiver configurations and introduces current-current feedback. Section 6.2 is focused on maximal bandwidth performance to illustrate the use of the transfer tools introduced in chapter [2–5]. Section 6.2 focuses on low-noise performance and illustrates the application of our noise tools, introduced in chapter 7–9.

Highlights

Highlights of this chapter on lightwave receivers using current-current feedback, originated from this study, are:

- *Third order* feedback loops over more than 500 MHz bandwidth [608].
- *Second order* feedback loops over more than 1 GHz bandwidth, in combination, with *noise tuning*
- *Low-noise performance* while using low-cost discrete components (3 pA/√Hz average noise over 1 GHz bandwidth using 1.5 pF photo diodes).

These receivers are applied to various lightwave transmission experiments, including the international projects ESPRIT 2054 (UCOL), RACE 2024 (BAF [108]) and internal projects including COSNET [107].

¹ The words *iterative synthesis* are use to distinguish a design approach from *trial-and-error* methods. An iterative synthesis is the improvement of a design in a structured way rather than random changes of circuits.

6.1. Lightwave receiver configurations

Lightwave receivers use one or more photo sensitive transducers such as PIN diodes, avalanche photo detectors (APD), photo-multipliers and camera-tubes. Some of these may be preceded by an optical amplifier for improving overall sensitivity. This chapter is limited to photo-sensitive devices in which the output current (photo current) is proportional to the lightwave signal level of interest. Two photo sensitive devices in shunt, as used in balanced receivers, can be treated as a single device. To simplify matters, all our examples use single PIN diodes as the photo sensitive device, although the concepts are more widely applicable. We will simply refer to all these devices as *photo-diode*. The pre-amplifier that amplifies the (electrical) signal current generated in the photo-diode is referred to as the receiver *front-end*.

For wideband lightwave receiver front-ends the basic design aspects are bandwidth, noise performance and large-signal behavior. Gain is of secondary importance since additional amplifiers may be used to increase the gain. The same applies for linearity considerations when digital signals are to be received.

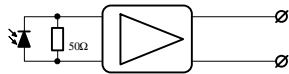
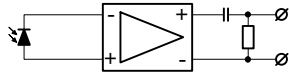
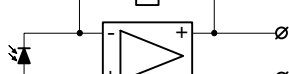
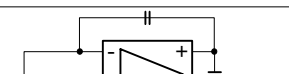
configuration	Band-width	noise	large signal
 (a) low impedance receiver	++	--	++
 (b) high impedance receiver	+/-	++	-
 (c) trans-impedance receiver	low R: ++ medium R: +/- high R: -	-- +/- ++	++ +/- -
 (d) current-current feedback receiver	++	++	++

Fig 6.17 A survey of various lightwave receiver front-ends

Numerous lightwave receiver concepts have been published in recent years. A survey of various receiver front-ends is shown in figure 6.1. This classification facilitates the assessment of limitations on the basic design.

This section 6.1 discusses these four classes in more detail, to illustrate the choice of the configuration shown in figure 6.1d for the design example. Section 6.2 and 6.3 discuss various design aspects concerning wide-bandwidth and low-noise performance.

6.1.1. Conventional receiver front-ends

Low impedance receivers. The most simple lightwave receiver configuration is shown in figure 6.1a. The photo current is accurately transformed into a voltage using a resistor, prior to amplification.

This configuration is advantageous in that it enables the use of long transmission lines between photo-diode and amplifier. The detector load and the amplifier input impedance must be matched to the characteristic impedance of the cable, usually $50\ \Omega$.

Another advantage is that any matched amplifier is applicable, since amplifiers with large bandwidth² and good large-signal performance are available.

The major disadvantage of low-impedance receivers is their poor noise performance. The thermal noise of the resistor usually exceeds the equivalent input noise of the amplifier, since it is as high as $18\ \text{pA}/\sqrt{\text{Hz}}$ at room temperature when using $50\ \Omega$ resistors.

High impedance receivers. Removal of all low value shunt resistors and high value series resistors in the amplifier reduces the overall thermal noise contribution significantly. This includes all resistors biasing the photo-diode and the input amplifier stage. Figure 6.1b illustrates the configuration. This approach leads to relatively high amplifier input impedance, the photo-current to input voltage transfer being dominated by the photo-diode capacitance. The transfer function, which is integrating in nature, is usually equalized by adding a differentiating network to the amplifier output (see figure 6.1b).

The advantage of this concept is to be found in its ultimate low-noise performance. Bandwidth is mainly limited by the output network used to equalize the transfer function. The major disadvantage is that the receiver can easily be overloaded by low-frequency signals. Of secondary importance is that harmonic distortion may exclude analog applications. This is because the transfer function is determined primarily by parasitic effects, and because the non-linear response of the transistor is not suppressed by feedback.

Transimpedance receivers. Some of the advantages of low- and high impedance receivers can be combined when using feedback [602]. The associated feedback configuration is shown in figure 6.1c. The transimpedance feedback loop controls the gain (and preferably also the bias) and transforms accurately the input current into an output voltage. The overall receiver performance is feedback dependent since the loop gain decreases with frequency in practical circuits. In figure 6.1c, three different feedback resistor values are considered: low, medium and high. They are:

- When the transimpedance resistance is chosen to be relatively low, its high thermal noise contribution will probably dominate the excess noise of the input stage. In this situation, the noise performance is poor as in the case of the low impedance amplifier. On the other hand, the feedback controls the gain over a wide bandwidth, since the loop gain is sufficiently high. As a result, this receiver has excellent bandwidth and large-signal performance.

² The corner frequency of the PIN-diode capacitance and the resistor may reduce the overall lightwave receiver bandwidth. For instance $1\ \text{pF}$ shunted with $50\ \Omega$ resistance and another $50\ \Omega$ input impedance limits the bandwidth to $6.4\ \text{GHz}$

- When the transimpedance resistance is chosen to be relatively high, its low thermal noise contribution is comparable to that of the high impedance amplifier. The loop gain is relatively low so that the feedback is limited to low frequencies. The transfer function and the large-signal behavior are comparable to that of a high-impedance receiver.
- When varying the transimpedance resistance from low to high, the overall performance changes between the extreme values as described above. For medium transimpedance values all three basic aspects are realized with moderate performance.

Most lightwave receivers are currently configured as transimpedance receivers.

Principal restrictions of conventional receiver frontends. Telecommunication systems require frequency independent response and high dynamic range. Feedback is an adequate method to improve the response function as well as large-signal performance. Therefore, a transimpedance amplifier, with frequency independent transimpedance, is commonly used.

In the case of passive feedback networks, a frequency independent transimpedance network inevitably requires (noisy) resistors. When relatively high transimpedances are used, the thermal noise of the feedback network can be made small compared to the noise of the active devices. High transimpedance values, however, are associated with high overall gain and this gain restricts the overall receiver bandwidth. On the other hand, lowering transimpedance values improves the bandwidth at the cost of noise performance, due to increased thermal noise of the feedback network. Therefore, bandwidth, noise and large-signal performance cannot be optimized simultaneously for the conventional receivers in figure 6.1a,b,c.

These conventional receivers have in common that they transform input current into output voltage. Their gain is therefore not dimensionless, but has the dimension of an impedance. This dimension forms a principal limitation when optimizing bandwidth, noise and large-signal performance simultaneously.

6.1.2. Improved receivers with current-current feedback

This study has considered in depth alternatives to the above and has resulted in current-current feedback amplifiers. We introduced this feedback principle for wideband lightwave receiver design in [608]. Current-current feedback combines the bandwidth and large-signal performance of transimpedance amplifier, with the noise performance of high-impedance amplifiers. This principle replaces the noisy resistive feedback network of a transimpedance amplifier by a low-noise network such as a capacitive current divider. A specific implementation is shown in figure 6.1d.

The feedback loop, can be used to flatten the current-current gain frequency response and the amplifier can handle relatively large input signals, similarly to transimpedance amplifiers.

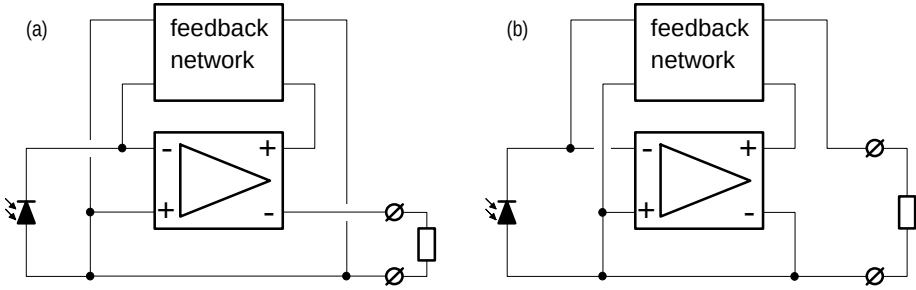


Fig 6.17 Generic block diagrams of lightwave receivers using current-current feedback.

Current-current feedback amplifiers require feedback networks, in shunt with the amplifier input and in series with the amplifier output. Figure 6.2 shows two block diagrams. The major difference is that in figure 6.2a the amplifier output port is floating to ground, while in figure 6.2b the feedback input port is floating.

The output signal is a current, which can be amplified with a second current-current stage or can be converted into a voltage using a load resistor.

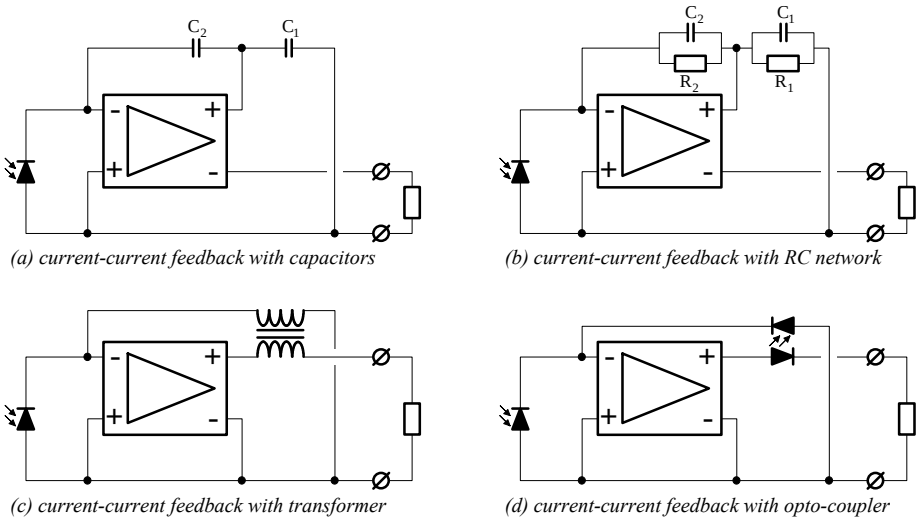


Fig 6.17 Specific implementation examples of lightwave receivers using current-current feedback

Figure 6.3 shows various specific implementation examples of lightwave receivers using current-current feedback.

- The current gain α of the pure capacitive variant in figure 6.3a is controlled by the ratio of the two capacitors, or more precisely by $\alpha = (C_1 + C_2) / (C_2)$.
- The RC-variant in figure 6.3b has some advantages to the configuration in figure 6.3a. It enables biasing of the input amplifier stage using the overall feedback loop. The resistors may be relatively high in value since the wideband feedback

performance is dominated by the two capacitors. Moreover, excessive voltages at low frequencies are avoided at the input node of the feedback network.

- The transformer variant in figure 6.3c eliminates the need for an amplifier with floating output terminals. The current gain equals the winding ratio of the transformer.
- The optical feedback variant, as is shown in figure 6.3d, circumvents the non-ideal performance of transformers while maintaining the advantage of non-floating amplifier terminals.

In the initial phase of this study, experiments were carried out using transformer current-current feedback. Bandwidths up to 100 MHz were shown to be feasible when transformers are used. Parasitic capacitances in the transformers limited the maximum bandwidth, and made further study uninteresting.

In the remainder of this text, design examples illustrating the implementation of current-current feedback amplifiers, using the RC-variant of figure 6.3b, will be considered.

6.1.3. Implementation of receivers using current-current feedback

The order of a feedback loop equals the dominant transfer order of the loop gain. The higher the feedback order is, the more complicated the design will be to control the stability of the loop. Therefore the loop order is in practice kept as low as possible.

In well-designed feedback amplifiers, each amplifier stage adds no more than one dominant pole to the loop while it provides maximum gain. In poorly designed feedback amplifiers, the gain per stage is significantly reduced in order to realize loops with low feedback order. This subsection discusses some well-designed feedback amplifiers having first-, second- and third-order feedback loops respectively.

First order loops. The simplest receivers with current-current feedback are similar to the first order configuration in figure 6.4a. A cascade of identical feedback amplifiers (figure 6.4b) increases the overall gain. Each loop consists of a single amplifier stage: two FET's in long-tailed pair configuration. This simple configuration has the advantage that stability compensation is not required.

On the other hand, the receiver performance is moderate because of the low loop-gain and because the amplifier stage must simultaneously match low-noise demands as well as output power requirements. We will pay no serious attention to these configurations.

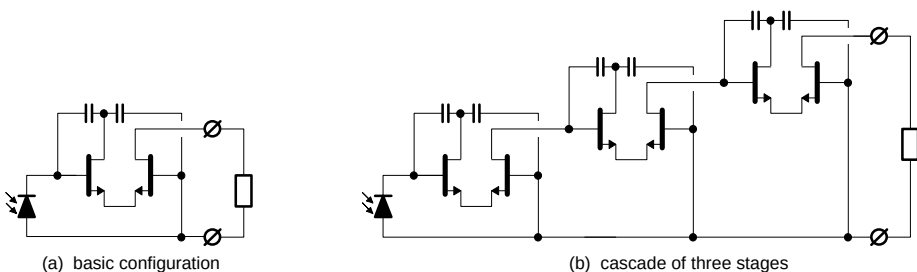


Fig 6.17 AC diagram of a simple lightwave receiver using first order current-current feedback.

Second order loops. Increasing loop gain by adding amplifier stages to the loop increases the available bandwidth and minimizes the noise contribution of the amplifier stages following the input stage. Figure 6.5 illustrates an implementation of overall current-current feedback over two amplifier stages. This study resulted in various receivers [608,108,607,609,610,612,613] using this configuration, with bandwidths in excess of 1 GHz.

A common-source configuration has been chosen for the first stage (T1), to maximize the loop-gain [406]. The design of T1 can be focused on optimal noise performance since the signal level to be delivered to the second stage is relatively low.

The second stage (T2+T3) is optimized for maximal output signal [406]. The current gain of this stage equals the current gain of transistor T2. Transistor T3 provides active decoupling³ from the signal voltage at the collector of T3. This voltage is maximal at frequencies below $\omega < 1/(R_1 \cdot C_1)$.

The output impedance of T3 is shunted to the input of the feedback network. Since T3 is configured in common base, its output impedance is relatively high and the deterioration of the feedback transfer by parasitic effects is relatively small. This property maximizes the overall loop-gain and makes T3 essential for taking full advantage of current-current feedback.

A phantom zero [406], as discussed in section 5.2, is used to controll the stability of the feedback loop,. This phantom zero is introduced by resistor R_{1s} that increases the feedback factor for frequencies above $\omega > 1/(R_{1s} \cdot C_1)$.

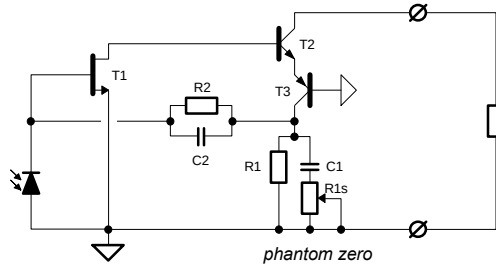


Fig 6.17 AC diagram of a lightwave receiver using second order current-current feedback. Resistor R_{1s} introduces a phantom zero to improve the stability of the overall feedback loop.

Third order loops. When a current gain greater than $100\times$ is required in combination with bandwidths of several hundreds of MHz, two amplifier stages are unable to provide the required loop gain. Figure 6.6 shows a lightwave receiver reported using third order current-current feedback. The first lightwave receiver, using third order current-current feedback over several hundreds of MHz (500MHz, $\alpha=100\times$), was developed in the course of this study [608].

³ Active decoupling is additional circuitry, like a buffer stage, for improving the isolation between output and input. Well-known examples are an inverting cascode stage (CE-stage, buffered by a CB-stage) and a non-inverting long-tailed pair (CC-stage buffered by a CB-stage). T2+T3 in figure 6.5 form a long-tailed pair. T4+T6 in figure 6.6 form a cascoded pair.

The input stage (T1) and output stage (T4+T5) are similar to the stages in figure 6.5. In this example, T6 has been added to the output stage to provide additional active decoupling of T4 from output signal variations at the collector of T6.

An intermediate stage (T2+T3) increases the loop gain. Transistor T2 provides the current gain, and transistor T3 provides the active decoupling of T2.

A phantom zero in combination with a pole-splitting network [406] are added to the loop to improve the overall loop stability. This compensation principle has been discussed in section 5.2.

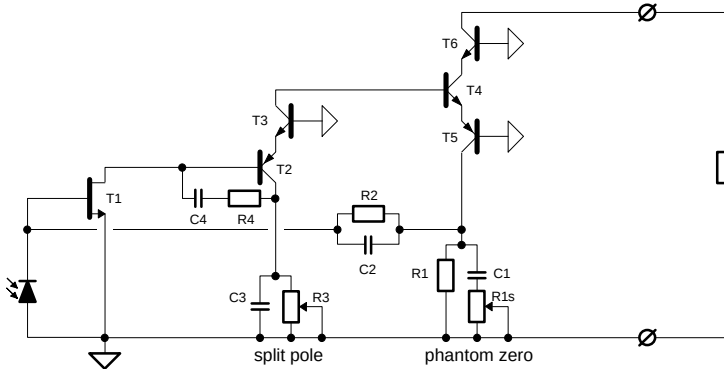


Fig 6.17 AC diagram of a lightwave receiver using third order current-current feedback. Resistor R_{1s} introduces a phantom zero to improve the stability of the feedback loop. The local feedback network with R_3 , C_3 , and R_4 , C_4 provides pole splitting for the same purpose.

So far, various implementation examples are shown of lightwave receivers using current-current feedback. An exact description of their implementation is beyond the scope of this text. More relevant for this text is the *strategy* that has been used to handle circuit design of higher order feedback loops. Section 6.2 is focused on this strategy.

6.2. Iterative synthesis focused on maximal bandwidth

An important aspect of high-performance lightwave receivers is wide bandwidth. The use of overall feedback in amplifiers, covering several amplifying stages, is preferable to circuit configurations using local feedback in various individual amplifier stages [406]. In general, overall feedback increases the available bandwidth and improves the overall linearity of the amplifier [406].

One may conclude that the number of amplifier stages in the loop should, therefore, be chosen as high as possible. Unfortunately, the transfer order of the loop-gain increases with the number of stages, and higher order feedback loops set stringent demands on control of the loop stability. The number of stages in an overall feedback loop must therefore be limited in practice.

Higher order feedback loops are commonly avoided in wideband applications. In the literature, state-of-the-art receivers have been reported achieving more than 20 GHz bandwidth, using local feedback or no feedback at all. When bandwidth is the only design parameter to be maximized, this may be the best approach. When aspects such as

cost, noise, linearity and large-signal performance are relevant, then higher order feedback loops are to be preferred.

This section 6.2 focusses on a *convergent* iterative synthesis technique for designing *higher* order feedback loops (order ≥ 2) exceeding 1 GHz in bandwidth. Design examples, realized using discrete components mounted on glass-epoxy printed circuit boards (see section 1.4.1) illustrate the practicality of the method.

6.2.1. Quantifying parasitic effects using circuit stripping techniques

In the initial stage of the study of lightwave receivers (1985), synthesis methods proposed by Nordholt [406] were used for designing the basic circuit. Reasonable results have been achieved up to 100 MHz on lightwave receivers using second order current-current feedback loops. Empirical extension to greater bandwidths using a circuit simulator failed. This was traced to a too limited understanding of the circuit parasitics. To facilitate a reliable simulation, a structured method has been developed for quantifying relevant parasitic effects. This subsection introduces a technique, termed *circuit stripping*, for locating inadequately modeled circuit elements and quantifying the associated error. It enables the extraction of improved models, as described in section 3.2, to adjust the overall circuit synthesis. The basic idea of circuit stripping is illustrated in figure 6.7.

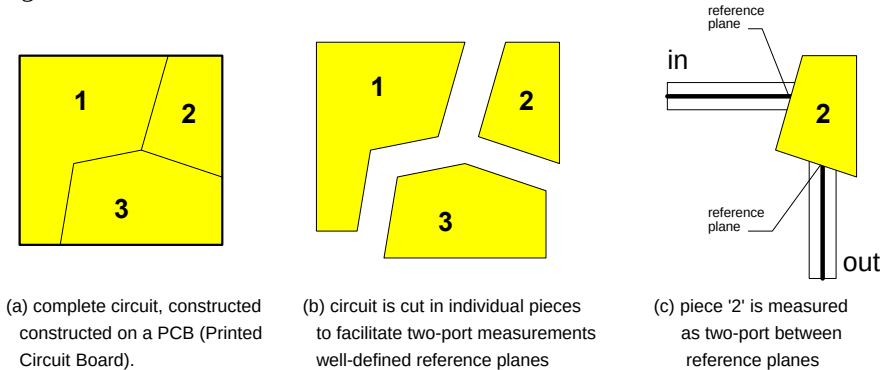


Fig 6.17 The basic idea of circuit stripping. The printed circuit board (PCB) has been cut into pieces to isolate individual sub-circuits. Each piece of PCB is extended with microstrips to facilitate full two-port measurements observed from well-defined reference planes.

Consider a circuit for which the simulation fails above a certain frequency. Figure 6.7a represents this circuit, constructed on a printed circuit board (PCB). The printed circuit board (PCB) has been cut in several pieces, as is shown in figure 6.7b, to isolate individual sub-circuits. The sub-circuits of interest are then to be extended with microstrip lines with equal and precisely known length, as is illustrated in figure 6.7c. This construction facilitates an exact positioning of reference planes for full two-port measurements, as described in section 3.1.3 and 3.2.1.

A practical implementation of the approach is illustrated in figure 6.8. The center of this photograph shows a lightwave receiver, using third order current-current feedback (the configuration of figure 6.6). The printed circuit board (PCB) has been cut in four pieces: the three individual amplifier stages and the overall feedback network. In figure 6.8, they are arranged around the receiver.

Full two-port measurements on each piece of PCB facilitate the 'reconstruction' of the complete circuit in a simulator. When all pieces are measured, the reconstructed overall transfer is adequately predicted. On the other hand, step-by-step substitution of each piece of PCB by an equivalent circuit model provides the means for identifying poorly modeled circuit elements. These steps are then repeated on a reduced scale until all suspected circuit elements have been singled out. The measured data may then be used to improve the associated model of these elements.

Using this circuit stripping technique, it became clear that the validity of commonly used hybrid- Π models for bipolar transistors is quite limited. The focus of the study was shifted in turn, albeit temporarily, to transistor characterization, resulting in improved BJT models as discussed in section 3.2.2.

Circuit-stripping of our lightwave receiver design have been proven in practice to detailed, quantitative insight into various parasitic effects. Some of them to be reproducible quantities and are used to control the overall gain. Other identified parasitic effects have been reduced by careful redesign of the PCB layout.

An example is capacitance C_2 in figure 6.5 and 6.6. The parasitic stray capacitance over R_2 is a reproducible quantity for a given PCB layout and serves the purpose of feedback capacitance C_2 .

Other examples are the parasitic effects originating from transistor T6 in figure 6.6. This transistor has been added to the circuit for providing adequate active decoupling of transistor T4 from signals at the collector output of T6. This is sometimes referred to as reducing the 'Miller' effect by a cascode configuration. Above a certain frequency, the benefits of T6 fade into insignificance and introduce new parasitics that are more harmful. As soon as this effect was identified and T6 was removed, the loop was compensable over a significantly wider frequency band.

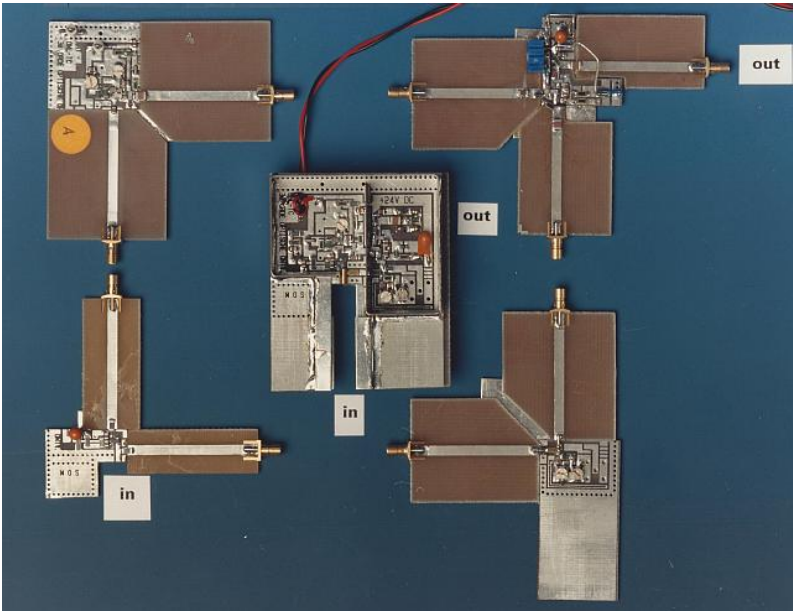


Fig 6.8 Example of circuit stripping. The printed circuit board (PCB) of this receiver has been cut into pieces to isolate the three amplifier stages and its overall feedback network. Each piece of PCB is extended with microstrips to facilitate transfer measurements observed from well-defined reference planes.

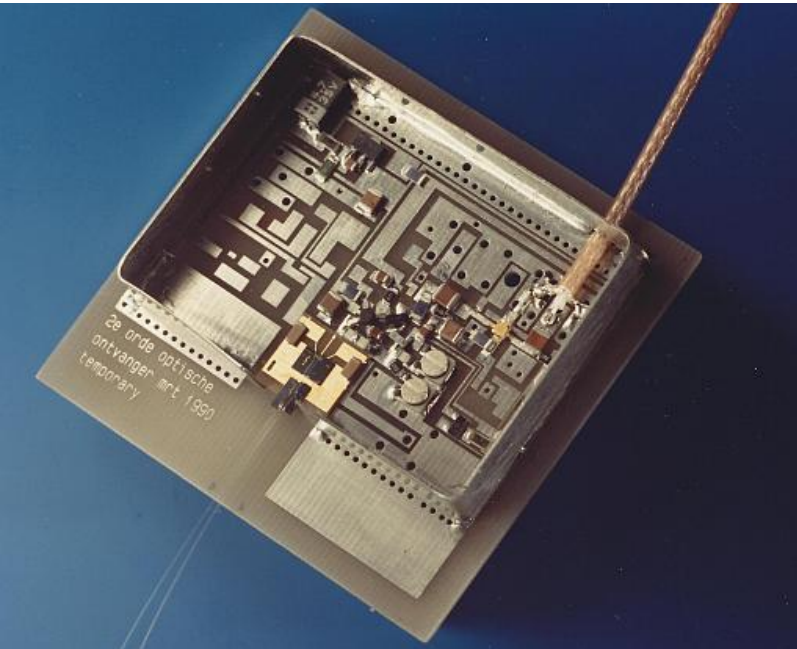


Fig 6.9 Lightwave receiver (balanced) using second order current-current feedback. All components being no part of the receiver pre-amplifier have been omitted.

6.2.2. Cumulative quantification of parasitics using pole-zero techniques

The *quantification* of individual parasitic effects, is merely one aspect of iterative synthesis. This facilitates reliable circuit simulation without distinguishing harmless parasitics from dominant parasitics. *Cumulative* quantification of the combined effect of various parasitics is required in order to whether particular parasitics are significant or not when combined. This cumulative assessment is essential because parasitic effects that are significant in specific sub-circuits or devices are not necessary of importance in the overall feedback loop.

In an integrated design environment, facilitated allowing acquisition and computer analysis of measured two-port parameters, designers are to quantify and cumulate parasitics with ease and in a structured way. Pole-zero extraction techniques, are eminently suitable for cumulative quantification parasitics from properly measured or simulated data. An experimental design environment, termed SABEL-CAE [111] and discussed in section 1.4.2, has been realized with many of the desired features.

In the future circuit simulators may support pole-zero techniques similar to that described herein. The designer simulates the circuit in two parts: a four-port feedback network and a two-port forward amplifier. The simulator evaluates the superposition model of the feedback circuit, as described in section 4.2, and extracts the dominant poles and zeros of the feedback loop. This is the desired cumulative quantification of parasitic effects.

When requested, the simulator provides the transfer function(s) of the compensation network(s), as described in chapter 5.

For an in-depth discussion of the associated algorithms see chapter 4 and 5. More effort has been spent on developing algorithms and software tools for pole-zero analysis and parameter extraction from measured data than was spent on the actual design of our lightwave receivers. Therefore, a short overview is given below to indicate the scale of this effort.

- From the very beginning the importance of measurements for improving the overall circuit simulation was evident. Therefore, data acquisition tools for network analyzers and spectrum analyzers have been developed to facilitate the required data processing. The associated software and hardware have been discussed in section 1.4.2, that refers to various internal reports.
- Poles and zeros are pre-eminently suitable for the analysis and synthesis of complex feedback loops. We started the development of a software tool for simulating root-loci of feedback loops. This program LOCUS has been discussed in various internal reports [113-118] and has provided the tools for developing the basic insight into the feedback theory of chapter 4 and 5.
- Pole-zero extraction is a convenient formalism for cumulative quantification of parasitic effects in measured data. Initially, a variant of the iterative Levenberg-Marquardt algorithm [410,114] has been developed for this purpose. The disadvantage of this algorithm is that starting values are required as input, and that it fails when providing inadequate starting values. Recently we developed linear pole-zero extraction methods as discussed in section 3.1.1. These algorithms require no starting values at all, are much simpler for implementation and have proved to be very robust.

- The pole-zero extraction algorithms facilitate an adequate pole-zero analysis without ignoring parasitic effects. Recently, we developed the missing links of the theory on feedback analysis and synthesis, as described in chapter 4 and 5, and implemented the associated algorithm for taking full advantage of this iterative synthesis.

The progress in performance of our receivers kept pace with the ever increasing performance of our tools for quantifying the cumulative effect of parasitics. An increasing number of parasitic effects have been identified using these tools, resulting in receivers performing in bandwidths as discussed in section 6.2.3.

6.2.3. Bandwidth performance of current-current feedback receivers

The circuit synthesis proposed by Nordholt [406] was used for the initial design of our lightwave receivers. An iterative synthesis improved various sub-circuits one by one. Circuit stripping as well as pole-zero extractions were used for this purpose:

- Some of the quantified parasitic effects are adequately *minimized* by proper redesign of the PCB layout and by adding active decoupling circuits. Some active decoupling circuits had to be removed, since they introduced annoying parasitics.
- The remaining parasitics were cumulated using the pole-zero extraction algorithms as discussed in chapter 4. Dominant parasitics are *neutralized* using adequate compensation networks as discussed in chapter 5.

Various high-performance receivers with third [608,607] and second order [609,610,612,613] feedback loops have been realized, using (parts of) this strategy. All these receivers are constructed using low-cost discrete (SMD) components on low-cost glass-epoxy printed circuit boards in microstrip print layout (see section 1.4.1).

Third order loops. The initial phase of our study was focused on third order feedback loops. The photograph in figure 6.8 on circuit stripping shows one of these prototype receivers. Receivers using third order feedback loops have proved [608] to be feasible for bandwidths exceeding 500 MHz and current gains of more than 100 $\times$. These receivers are constructed from discrete components including photo diodes with 1.5~2.0pF photo diode capacitance.

Second order loops. Further increase of bandwidth has not been achieved with third order feedback receivers. The stability compensation was the limiting factor. Various simplified receivers using second order feedback loops have proven feasible for bandwidths of more than 1 GHz and current gains of more than 27 $\times$.

The photograph in figure 6.9 shows an implementation of a receiver, using the circuit configuration shown in figure 6.5. Figure 6.10 shows the associated transfer function [609,610,612]. This receiver was configured as a balanced receiver with 1.5 pF total photo-diode capacitance [612], and applied in transmission links for an international ESPRIT 2054 project (UCOL). Similar lightwave receivers have been developed for ATM transmission experiments [108] in an international RACE 2024 project (BAF) or used in various other coherent optical systems [107] (COSNET).

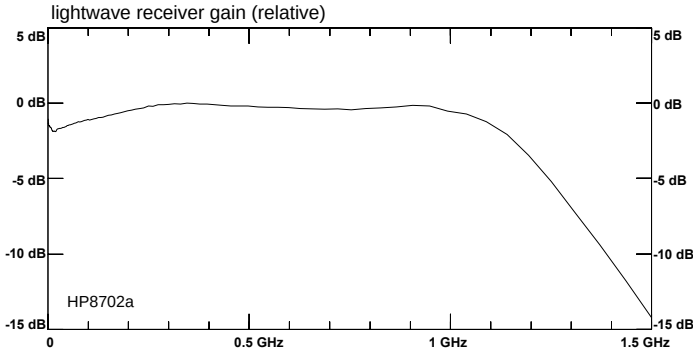


Fig 6.17 Measured transfer of a lightwave receiver using second order current-current feedback. This receiver, shown in figure 6.5, was configured as a balanced receiver with 1.5 pF total photo-diode capacitance.

The ultimate bandwidth performance. The results discussed in this section are indicative for how far an iterative synthesis using sophisticated tools can push the overall receiver performance towards physical limits when the technology is restricted to low-cost technology.

In our design example, the actual adjustment of compensation networks was inevitably performed on a trial-and-error base using a circuit simulator. This conventional approach was initially inevitable because our feedback synthesis techniques of section 5.3 were still under development at that time. Therefore it is plausible that the ultimate performance limit has not been reached.

Experimental receivers can be build with bandwidths of more than 2 GHz, although the pass-band ripple of these receivers is too high. Therefore it is plausible that bandwidths of more than 1 GHz are feasible when *all* techniques of chapter 5 are used to support the iterative synthesis of feedback circuits. Further improvements to receiver bandwidth are still undergoing study.

6.3. Iterative synthesis focused on minimal noise

The design of our application example has focused on transfer aspects of the iterative synthesis. It illustrates the value of the transfer-oriented part of our study, as discussed in chapter 4 and 5. This section 6.3 focuses on various noise aspects of iterative synthesis, as introduction to the discussions on noise in chapter 7, 8 and 9.

Wide bandwidth as well as low-noise performance are important aspect of lightwave receivers. One of the major advantages of receivers using current-current feedback is that the associated feedback network is nearly loss-free and therefore almost noise-free. It means that receiver design focused on maximum bandwidth does not seriously deteriorate the overall noise performance. Moreover, the use of *overall* feedback over several amplifier stages improves the overall noise performance. This is because the gain of a first amplifier stage is maximized by avoiding local feedback, and this minimizes the noise contribution of all succeeding amplifier stages.

This section discusses various design aspects of minimizing noise in lightwave receivers by tuning. Again, it emphasizes the importance of circuit stripping techniques and pole-zero techniques when tuning networks are added to the loop, to handle noise tuning as well as feedback stability. The importance of noise characterization as part of an iterative synthesis is discussed in section 6.3.5.

6.3.1. Noise optimization in the absence of tuning

The noise performance of well-designed receivers using overall current-current feedback is in practice determined by the noise of the first amplifying stage. Minimization of equivalent input noise is therefore mainly focused on the first stage.

Figure 6.11 shows an equivalent model of a noisy two-port (first amplifier stage) that amplifies the signal generated by a current source (photo-diode). The two equivalent noise sources are in principle correlated and frequency dependent sources. This two-port noise model is discussed in section 7.3.1. BJT's as well as FET's are suitable for use as the active device in this amplifying two-port.

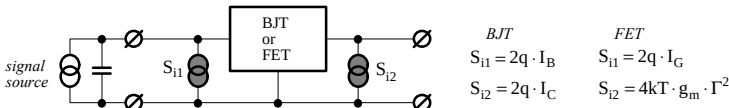


Fig 6.17 Simplified noise model of the input stage of an amplifier. The two equivalent noise sources are in principle correlated and frequency dependent. The formulas in this figure represent a first order approximation related to the collector bias currents I_C , the base bias currents I_B , the gate leakage current I_G and the transconductance g_m . The factor Γ is an empirical correction factor: $\Gamma \approx 1$.

FET's in common source configuration or BJT's in common emitter configuration provide maximal gain and optimal low-noise performance [406]. In these configurations the noise intensity spectra S_{11} and S_{12} are nearly white and almost uncorrelated [406]. In a first order approximation S_{11} equals the shot noise of the base bias-current or gate leakage current, while S_{12} equals the shot noise of the collector bias-current or the thermal channel noise [406]. In addition, thermal noise of bias resistors must be added to S_{11} and S_{12} .

The current gain α of the amplifying two-port reduces the contribution of S_{12} to the total equivalent input noise. In wideband applications, this current gain $\alpha \approx -(j \cdot f) / (f_T)$ decreases with the frequency as specified by the effective transition frequency f_T of the amplifier stage⁴. Therefore the total equivalent input noise approximates:

$$S_i \approx S_{11} + S_{12} / |\alpha|^2 \approx S_{11} + S_{12} \cdot |f / f_T|^2$$

Noise measurements, as shown in curve *a* of figure 6.17, demonstrate that the measured noise current ($\sqrt{S_i}$) increases almost linearly with the frequency, as predicted by this

⁴ Note that the effective transition frequency of the *input stage* is less than the transition frequency of the individual active devices. Its value is lowered by the photo diode capacitance and all other parasitic capacitances at the input. This holds especially when FET's are used because the transconductance of FET's is usually significantly lower than the transconductance of BJT's with equal transition frequency.

formula. This approximation often holds for FET's as well as for BJT's in wideband lightwave receivers. At relatively low frequencies the current gain α of FET's is extremely high, which makes them in these cases superior in noise performance to BJT's. At relatively high frequencies this preference is not so obvious since in these situations the influence of ω_T is dominant.

Noise optimization is primarily a matter of finding an optimal combination of f_T , S_{11} and S_{12} by selecting an optimal active device and optimal bias conditions [606]. Nordholt [406] has discussed this general procedure in detail from which we concluded that FET's are preferable to BJT's in our design example because of their low S_{11} .

The transition frequency f_T is an important parameter for low-noise design, and parasitic capacitances such as photo diode capacitance easily deteriorate this transition frequency. This holds especially when FET's are used in combination with diode capacitances as high as 1.5 pF. Specifically in this case, the circuit stripping techniques of section 6.2.1 and the circuit model extraction algorithms of section 3.1 and 3.2 are of primary importance for optimizing the noise performance of the amplifiers.

6.3.2. Noise tuning in absence of overall feedback

A more complex manner of improving the overall noise performance may be found in increasing the current gain α of the first amplifier stage using resonance techniques [601,603,604]. Figure 6.12 illustrates various tuning networks for connecting signal current sources (photo diode) to the input stage of receiver front-ends. We may distinguish the following tuning examples:

- *Shunt tuning*, as shown in figure 6.12a, uses a band-pass transfer function and cancels the capacitive diode- and input impedance within a small frequency band. Below this band the signal is shorted to ground making this type of tuning unsuitable for wideband applications extending relatively low frequencies.
- *Series tuning*, as shown in figure 6.12b, uses a low-pass transfer function, peaking in the vicinity of a corner frequency. When this frequency is equal to the bandwidth of interest, series tuning improves the noise performance in the high end of the pass band and deteriorates it above this band. Noise measurements, as shown in curve *a* in figure 6.17, illustrate how effective this approach can be.
- *Higher order tuning networks*, as shown in figure 6.12c, may, by introducing additional degrees of freedom, extend the tuning bandwidth when properly adjusted. The illustration depicts a ladder configuration, although various other combinations of inductors and capacitors are applicable.
- *Transformer tuning*, as shown in figure 6.12d, is just another version of higher order tuning.

These examples are all restricted to lumped elements. In practice the values of the tuning inductors are very low (e.g. below 50 nH for 1 GHz applications). Distributed tuning networks, such as short transmission lines, stubs and dielectric tuning tabs (e.g. $\epsilon_r > 100$), are therefore viable alternatives.

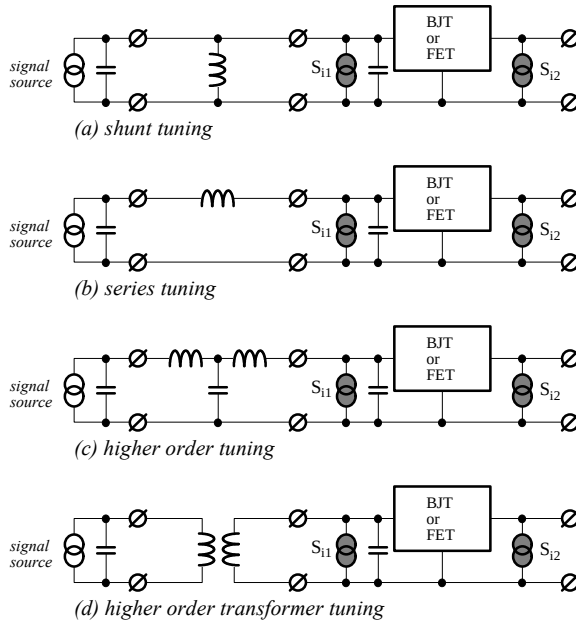


Fig 6.17 Various examples of tuning networks for improving the overall noise performance.

At the course of this study, experiments have been carried out with a variety of front-end receiver and tuning configurations. For an overview of various tuning configurations, see Jacobsen and Kan [603,604] or van der Vleut [611]. Figure 6.13 and 6.14 show a series tuned front-end, without feedback, using a transmission line and an additional dielectric tuning tab [611]. Figure 6.15 shows a front-end without feedback using higher order tuning networks.

Parasitic effects play a dominant role in tuned front-ends, and a tuning network consists in part of parasitic elements. Moreover, most elements of a tuning network cannot be adjusted when the actual value differ from the desired value. This emphasizes the importance of precise control of these parasitic elements to make them reproducible and accurately known.

The parasitic effects of the tuning networks have been quantified by various techniques as described in chapter 2, so that iterative synthesis may be effectively applied.

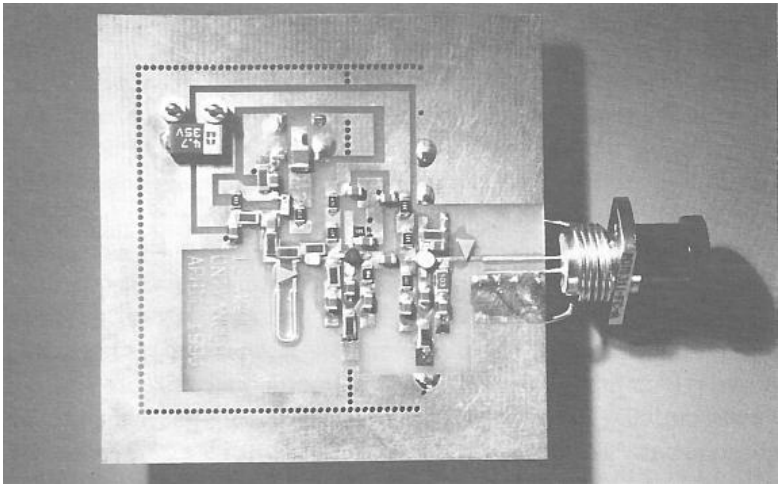


Fig 6.13 Tuned front-end (without feedback) using a transmission line and a dielectric tab as tuning element.

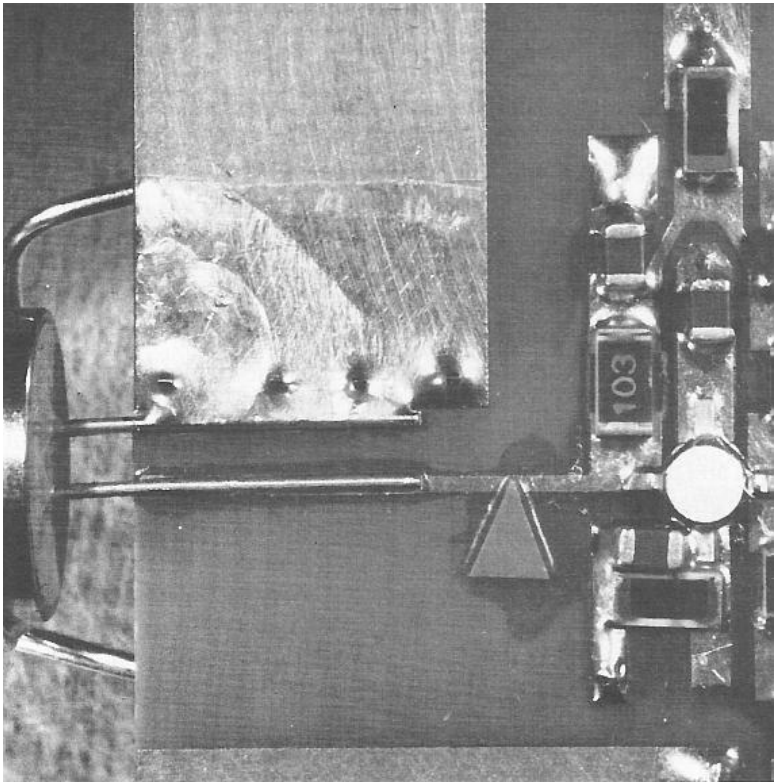


Fig 6.14 Close-up of the tuned receiver shown in figure 6.13.

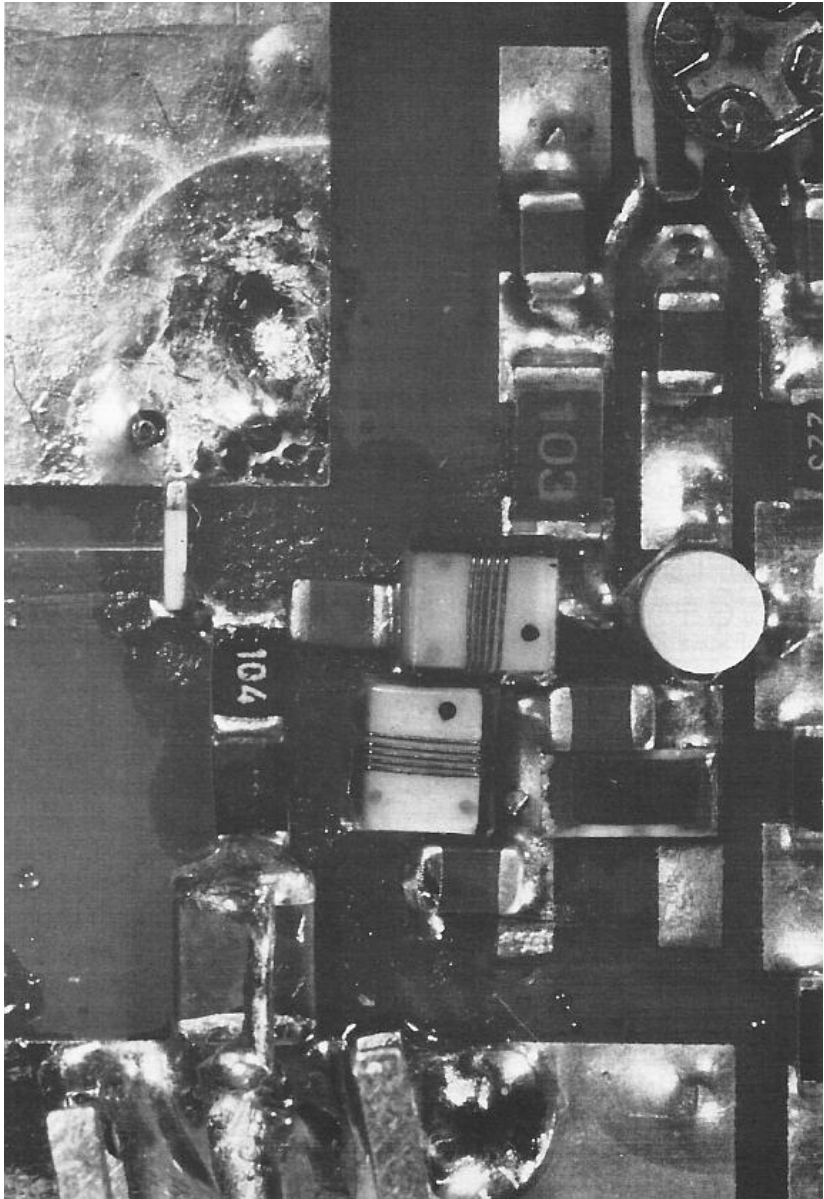


Fig 6.15 Close-up of a tuned front-end (without feedback) using two wound inductors and a parasitic third inductor as tuning element.

6.3.3. Noise tuning in combination with overall feedback

Tuning networks improve noise performance by modifying the current gain of the first amplifier stage. The principle drawback of noise tuning is that it deteriorates the transfer function of the lightwave receiver. External equalizers are required to eliminate this effect.

Overall feedback loops can make the use of equalizers superfluous. On the other hand, tuning networks affect the loop-gain and may endanger the overall stability of the feedback loop. In practice, overall feedback loops in combination with input tuning are commonly avoided.

This study has demonstrated [613] the viability of using higher order feedback in *combination* with input tuning. Tuned receivers have been realized using second order current-current feedback loops with bandwidths up to 900 MHz. Their noise performance is discussed in section 6.3.4.

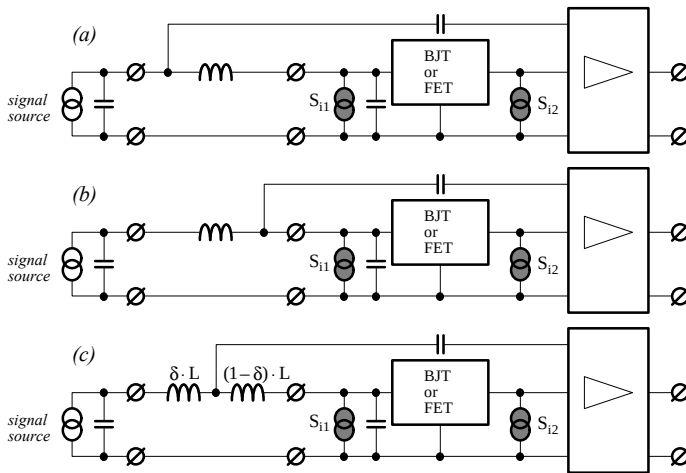


Fig 6.17 Typical examples incorporating input tuning in the feedback loop to improve the overall noise performance in lightwave receivers. When example c is well-designed the feedback current is frequency independently distributed over the source impedance and the input impedance.

In wideband design, the manner of interconnecting the feedback networks to the tuned input is the determinative factor in establishing feasibility. Figure 6.16 shows three typical examples. Configuration *c* is to be preferred [613] in wideband feedback applications for the following reasons:

- Figure 6.16a shows a configuration in which the serial tuning inductor is placed within the feedback loop. In case of infinite loop-gain, the overall transfer (output signal / photo current) is fully controlled by feedback and the tuning inductor will not deteriorate the overall transfer function. This is an important advantage. In practice, however, the tuning inductance plays a significant role at frequencies close to the high end of the receiver pass band. The loop-gain is then relatively low, and the tuning network inevitably degrades the overall transfer. Furthermore, in

most wideband applications the loop stability cannot be compensated. This is a serious disadvantage.

- In figure 6.16b the serial tuning inductor is placed outside the feedback loop. In the case of ideal photo-diodes with infinite output impedance, the loop-gain is not affected and the serial tuning inductor does not endanger the stability of the feedback loop.

In practice, the photo-diode shunt capacitance in combination with the tuning inductor lowers the loop-gain. The loop stability cannot then be compensated in most wideband applications. In this situation, the transfer function as well as the loop-gain are deteriorated by the tuning network.

- Figure 6.16c depicts a superior solution that eliminates the loop-gain deterioration. When properly adjusted, the feedback current is frequency independently distributed over photo-diode capacitance c_d and input capacitance c_g of the input stage. In the case that $c_d=c_g$, the tuning inductors $\delta \cdot L$ and $(1-\delta) \cdot L$ are equal. This alternative solution enables stable operation of higher order feedback loops in combination with input tuning.

This discussion is indicative of the complexity involved in using higher order feedback loops in combination with noise tuning in wideband applications. Parasitic effects play a dominant role. *Iterative* synthesis is required to facilitate stable feedback operation as well as adequate noise tuning.

In the design of the lightwave receivers described in this text, circuit parameters for adequate circuit simulation are extracted from circuit stripping techniques. The use of root-locus and pole-zero tools, as implemented in our SABEL-CAE [111] design environment, facilitated the design of wideband higher-order feedback loops.

6.3.4. Noise performance of current-current feedback receivers

The measurements in figure 6.17 show the equivalent input noise achieved for our lightwave receiver design examples. It represents the noise of lightwave receivers [613] using second order current-current feedback, a balanced pair of photo-diodes (1.5 pF) and an FET as input stage.

Curve *a* depicts the equivalent noise current ($=\sqrt{S_i}$) of a receiver front-end without input tuning. This measured curve is superimposed on a curve: $S_i=S_{i1}+(f/f_T)^2 \cdot S_{i2}$, as explained in section 6.3.1. The quantities S_{i1} , S_{i2} and f_T are parameters of the associated noise and transfer model. Curve *b* represents the noise current of a similar receiver using series tuning.

The average noise of the tuned front-end equals 3 pA/ $\sqrt{\text{Hz}}$. This noise level is as low as the thermal noise current of a 1.8 k Ω resistor or as the shot noise associated with a dc-current of 28 μA . These results have been achieved with low-cost components photo-diode capacitance in excess of 1.5pF. Comparable state-of-the-art receivers have been described using superior higher-cost components such as photo-diodes with lower shunt capacitance and avalanche photo diodes (APD's). This bodes well for the use of sophisticated design strategies to push overall receiver performance towards physical limits.

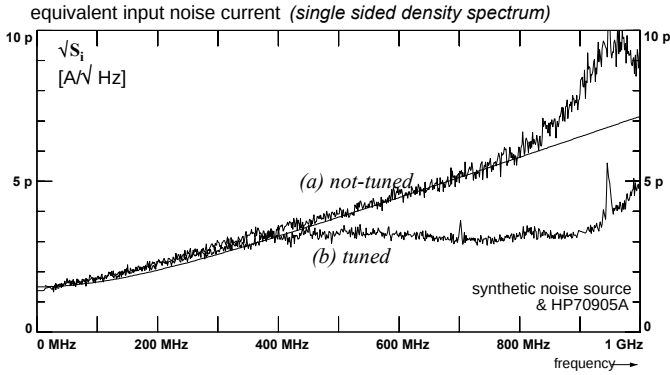


Fig 6.17 Measured input noise of two lightwave receivers [613], using the measurement techniques discussed in section [8.4] It represents the noise observed from an electrical reference plane, between photo-diode and amplifier.

Reliability of noise simulations

The prediction of noise in receivers without tuning is reasonably succesful when simple noise models are used in combination with transfer parameters extracted from measurements. For instance, the minimum noise level ($\sqrt{S_{11}}=1.5 \text{ pA}/\sqrt{\text{Hz}}$) is restricted by the thermal noise of the gate bias resistor⁵ in the feedback network and by the gate leakage current of the input FET. Noise increases with frequency, may be predicted by taking into account the total input capacitance (photo-diode and FET) and the FET transconductance⁶. The noise peak at the high end of the pass band probably originates from a parasitic tuning effect, that is to be identified.

Reliable prediction of the noise in receivers with tuning requires more sophisticated models. The tuning dip in figure 6.17 is not as deep as expected based on noise calculations using simple models. One of the reasons is that thermal noise effects in lossy tuning inductors have not been taken into account. Another reason is that *estimated* FET models were used, which probably requires further refinement.

6.3.5. The role of noise measurements in an iterative synthesis

Most of section 6.3 is focused on low-noise design using *measured* transfer parameters (circuit stripping) in combination with *estimated* noise parameters (the formulas in figure 6.11). This is because adequate instruments for transfer measurements on microwave and lightwave receivers are commercially available. Currently (1994), this does not hold for optical noise measurements on lightwave receivers, and is only partly true for electrical noise measurements⁷.

⁵ The thermal noise of a 10k Ω resistor is 1.3 pA/ $\sqrt{\text{Hz}}$.

⁶ The parameters, reconstructed from noise and transfer measurements are: $S_{11}=1.5\text{pA}/\sqrt{\text{Hz}}$, $S_{12}=34\text{pA}/\sqrt{\text{Hz}}$, $f_T=5.2\text{GHz}$, $C_{tot}=2.5\text{pF}$ (diode=1.5pF, FET=0.8pF, stray=0.2pF), $g_m=82\text{mA}/\text{V}$.

⁷ We note that instruments for full two-port noise measurements are commercially available. They focus, however, on microwave applications.

It is obvious that designing receivers based on estimated noise parameters yield sub-optimal results. Therefore, an iterative synthesis must also be based on *measured* noise parameters. The basic aspects are discussed below.

Noise measurements. Measurement of overall receiver noise is required to validate the noise performance predicted by estimated noise models. In simple situations, such as receivers without tuning, this approach has proven adequate. It provides some understanding of parasitic noise effects required for optimal low-noise design.

When tuning networks are involved, our measurements of overall receiver noise indicate a deviation between predicted and measured noise. Two-port noise measurements on individual devices, such as the input FET, are required to quantify parasitic noise effects that are inadequately modeled or that not have been taken into account. This would supplement our understanding of parasitic noise effects and enable further improvement of front-end design. Up to 5 dB additional improvement in wideband noise level is expected [613] for our design examples when the theoretical optimum has been achieved with similar components

Noise simulations. Simple⁸ noise simulations have proven adequate when designing optimal low-noise receivers in the absence of noise tuning. Thermal noise effects due to lossy tuning inductors were not taken into account in our simulation of tuned receivers. Further, the noise in the input FET was assumed to be white and uncorrelated.

When the simulation is based on reliable noise models for FET's and BJT's, extracted from two-port noise measurements, then these sources are probably correlated. Therefore more sophisticated noise simulations are required for further improvement of tuned receiver noise.

New developments. The second part of this study (chapter 7-9) is focused on algorithms for improved noise simulation, algorithms for noise parameter extraction of measured data, and on techniques for noise measurements

Chapter 7 discusses a general theory on noise, to facilitate sophisticated noise simulations of arbitrary linear circuits. It illustrates the extraction of thermal noise effects from transfer measurements.

Chapter 8 introduces various novel noise measurement techniques. It illustrates the measurement of electrical noise of *optical* receiver and full two-port noise characterization of semiconductor devices.

Chapter 9 describes the development of a novel measurement instrument to facilitate the desired noise measurements. This lightwave instrument generates white noise over tens of GHz.

⁸ Simple noise calculations use AC transfer analysis for a finite number of uncorrelated white noise sources. Circuit simulators supporting this method calculate the RMS sum of all white noise sources at a specified output node. This value is then divided by the gain to calculate the equivalent input noise. Pspice® [128] is an example of a circuit simulator using this approach. Modern microwave simulators use a more complete noise algorithm.

6.4. Conclusions

The application of the tools of this study in the support of iterative circuit synthesis for lightwave receiver design has been discussed. First cut receiver design was improved by extracting various circuit parameters from measured data. The extracted values provide a deeper understanding of parasitic effects and enable a designer using iterative synthesis to obtain optimal circuit performance.

Receiver performance

An adequate classification method was used to select an optimal lightwave receiver configuration based on current-current feedback. The examples make use of higher order feedback loops, demonstrating the applicability of higher order feedback in broad band design. Using iterative synthesis techniques, we developed the first lightwave receiver with third order current-current feedback over more than 500 MHz bandwidth [608]. The feasibility of second order feedback loops in combination with noise-tuning networks has been demonstrated over more than 1 GHz in bandwidth. An average noise level of 3 pA/√Hz has been achieved over this frequency band in combination with photo diodes having 1.5 pF shunt capacitance.

This performance is indicative of the ability of iterative synthesis using state-of-the-art tools to push overall receiver performance towards physical limits, even when low-cost technology is mandated. It is reasonable to expect that when all the tools of this study are applied, greater bandwidths will prove feasible.

All of the design examples of this chapter have been realized using discrete components mounted on low-cost glass-epoxy prints (see section 1.4.1). Although neither integrated circuit solutions thin film solutions have been investigated as yet, the application of the methods described in this book should prove applicable. Similar conclusions hold for other technologies, such as chips, although on a different (frequency) scale.

iterative synthesis focused on maximal bandwidth

Section 6.2 was focused on feedback transfer aspects of the design example. This study resulted in the development of *circuit stripping* techniques to quantify parasitic effects in the feedback loop. Circuit stripping is a structured method for locating inadequately modeled circuit elements. It enables the extraction of realistic model parameters from measurements, as described in chapter 2.

Parasitic effects that are significant in sub-circuits and devices are not necessarily significant in the overall feedback loop. This study resulted in various *pole-zero techniques* to cumulate the effect of various parasitics in the overall feedback loop. It enables the extraction of dominant poles and zeros in the loop, as described in chapter 3. The dominant poles and zeros have been compensated by adequate compensation networks, as described in chapter 5. Design of the compensation network was initially performed on a trial-and-error base using a circuit simulator. New *compensation algorithms*, introduced in chapter 5, facilitate the structured design of the compensation networks as well as the devising of new compensation concepts.

iterative synthesis focused on minimal noise

Section 6.3 focused on noise aspects of the design example. Detailed knowledge of various parasitic effects enables the designer of lightwave receivers to combine noise tuning networks with higher order feedback loops.

Simple noise simulations on tuned receivers have been shown to be only moderately reliable. This was because thermal noise effects in the tuning elements and noise correlation in the active devices are not taken into account. The *general noise theory* in chapter 7 facilitates a more reliable simulation and in principle enables the design of improved tuning networks.

Noise models with estimated parameters were used in the noise simulation. The novel measurement techniques of chapter 8 to measure the overall *receiver noise* facilitate the identification of relevant parasitic noise effects. Noise models may be refined and tuning elements adjusted to improve the low-noise performance of the receiver. Ultimate performance requires full *two-port noise measurements* on the individual devices to insure reliable noise simulations. Associated measurement techniques are discussed in chapter 8, using a synthetic noise generator as described in chapter 9.

Chapter 7

Formal description of noise

The design of high performance wideband feedback amplifiers requires optimization of transfer as well as noise performance. Nordholt [406] developed a design synthesis that handles either design goal. The preceding chapters have dealt with Nordholt's synthesis to feedback amplifier transfer in the case that active decoupling circuitry is ineffective in reducing parasitic effects. The proposed improvements rely extensively on transfer measurements between well-defined reference planes and on the extraction of adequate device models.

State-of-the-art noise measurements are not as accurate and straight forward as state of the art transfer measurements. The application examples in the previous chapter have, therefore, relied on *estimated* transistor noise models. The use of estimated noise models yields in many cases reasonable results. Nevertheless, the design of amplifiers with optimal noise performance requires more accurate models. Accurate noise measurements on lightwave receivers have indicated that simulated and measured noise are poorly correlated. This difference is relevant, for instance, when designing lightwave receivers with *optimal* wideband noise tuning.

Improving the design of low noise amplifiers requires the use of improved noise models extracted from *measured* data. Furthermore, simulation techniques for adequately handling the thermal noise of passive components with parasitic effects, such as lossy inductors in input-tuning networks (see section 6.3.2), are required.

Chapter 7 reviews the mathematical tools for (1) analyzing and measuring noisy circuits and (2) extracting device noise models. Measurement techniques will be discussed in chapter 8 and 9.

State of the art

Many articles have been published on modeling and characterizing semiconductor devices. Microwave applications are often focused on minimizing noise using resonance techniques. The minimization of noise figure in a small frequency band is a prime goal. From this practice, *spot noise parameters* have been evolved, as is described in section 7.3.2. These parameters are dedicated to narrowband noise optimization, however, they are very inconvenient when designing wideband amplifiers. As a result, most publications are focused on spot noise parameters. The same applies to commercially available microwave circuit simulators such as Touchstone[®][124], which often interfere with gaining insight into the noise performance of wideband amplifiers.

Measured noise parameters of semiconductor devices are at best sparingly available. An increasing number of manufacturers specify their semiconductor devices in terms of spot noise parameters, however, they are usually restricted to a few frequencies above 1 GHz. These frequency intervals chosen are often too coarse for adequate verification of device noise models.

On the other hand, various models have been published that rely on physical understanding of the devices. Some authors have compared their own models with measured data. These measurements indicate a rough agreement between model and measured noise, and are often restricted to a few frequencies above 1 GHz. When adequate noise measurements at several frequency points below 1 GHz are lacking, there is no way of judging whether parasitic noise effects are of relevance¹ to wideband applications or not.

Highlights

This chapter starts with unambiguous definitions of various signal and noise spectra. Most of the chapter relies on known² concepts. The highlights of this chapter, as considered in this study, are:

- Detailed and compact survey of known matrix methods for describing noisy multi-port networks. Voltage-, current- and wave methods are discussed simultaneously, and the wave methods include waves normalized to arbitrary (complex) reference impedances.
- Detailed and compact summary of various two-port noise parameters.
- Generalization of transformation rules between spot noise parameters and the wave correlation matrix. Their validation has been extended to waves normalized to reference impedances with negative real part.
- Introduction of a promising new concept for specifying the noise of semiconductor devices: *autonomous noise parameters*. They may simplify the extraction of transistor noise models from two-port noise measurements.

This chapter forms the theoretical foundation for the noise measurements that are described in chapter 8 and 9.

¹ An example of a parasitic effect of minor importance is induced gate noise in FETs. Nordholt [406] has shown that this effect is of secondary importance in the case of wideband amplifiers. Since adequate measurement data is lacking, it is conceivable that these parasitic effects are small compared to others that have been ignored. This illustrates the importance of using device noise models that are entirely based on measured data.

² We emphasize that this chapter reviews in essence known microwave concepts. Nevertheless, any summary perusal of the plethora of microwave texts should be sufficient to illustrate the fact that in general these texts survey an incomplete overview of the subject, with restricted validation of the formulae. Some texts use differing definitions among confusing typographical errors. Many textbooks and articles had to be consulted to complete this compact overview.

7.1. Definitions of signal and noise spectra

Section 7.1 considers unambiguous definitions for various signal spectra, as used in this study. This is not as obvious as it might seem. Most article and textbook discussions omit these definitions and use the letter *S* for some spectral quantity with a variety of names like (1) power spectrum, (2) spectral power density, (3) power density spectrum, (4) energy density spectrum, (5) spectral density, (6) spectral intensity or (7) spectral noise power. Nevertheless, there is no commonly accepted definition whether the single sided variant (S_{ω}) is meant or the double sided variant (D_{ω}).

<i>single</i> (0...∞) <i>4kTR</i>	<i>double</i> (−∞...∞) <i>2kTR</i>	<i>reference</i>
-	$2\pi \cdot W$	<i>Haus et.al. "Representation of noise in linear twoports"</i> Proceedings of the IRE, jan 1960, p69-74
S	-	<i>Ziel "Noise: Sources, Characterization, Measurement"</i> , Prentice-Hall, Englewood Cliffs New Jersey 1970
-	S	<i>Hullett, Muoi "A feedback receiver amplifier for Optical Transmission Systems"</i> IEEE Trans. on Communications, oct 1976, p 1180-1185
-	S	<i>Smith, Hooper, Garrett "Receivers for optical comm: a comparison of avalanche photodiodes with PIN-FET hybrids"</i> Opt. and Quantum Elect. 10 (1978) p293-300
-	S	<i>Davidse</i> Electronic amplifiers and phaselocked loops 1982. VSSD-Delft, the Netherlands (in Dutch).
S	-	<i>Nordholt "Design of high performance negative feedback amplifiers"</i> , Elseviers, Amsterdam, 1983, ISBN 0-444-42140-8
-	S	<i>Jain, Kumar, Gupta "Design of an optimum optical receiver"</i> J. of Optical Communications, 6 (1985) 3, p106-112
S	-	<i>Gupta et.al. "MW Noise characterization of GaAs MESFETS: Evaluation by ..."</i> IEEE transactions on MTT, vol 35, no 12 dec 1987, p1208-1217
-	S	<i>Howard, Jeffery, Hullett "On the noise of high-transimpedance amplifiers for long-wavelength pulse OTDRs"</i> Optical and quantum electronics 19 (1987) p123-129
S	-	<i>Darcie, Kasper, Talman, Burrus "Resonant pin-FET receivers for lightwave subcarier systems"</i> J. of Lichtwave technology vol 6, no 4, april 1988, p582-589
-	S	<i>Jacobsen, Kan, Garrett "Improved design of tuned optical receivers"</i> Electronic letters, vol 23, no15, july 1987, p787-788
-	S	<i>Einarsson "Error probability of optical receivers"</i> J. of Optical Communications, 11 (1990) 4, p137-141
S	-	<i>Schneider "Reduction of spectral Noise Density in pin-HEMT lightwave receivers"</i> J Journal of lightwave Techn. vol 9, no 7, july 1991, p887-892
-	S	<i>Dobrowolski</i> Introduction to computer methods for microwave circuit analysis and design. Artech House, Boston 1991, ISBN 0-89006-505-5

Fig 7.1 A survey of various publications related to intensity spectra. The symbol S is often used, however, there is no commonly accepted convention whether the single or the double sided spectrum is meant. This illustrates the importance of defining this quantity to avoid confusion when noise spectra are discussed.

To illustrate how confusing common practice is, figure 7.1 surveys various publications that use the letter *S* for intensity spectra, some of them indicating single sided spectra,

and others double sided spectra. Quite clearly, it is meaningless to specify noise levels S of amplifiers when an unambiguous definition of S is lacking. This is because *single* sided and *double* sided spectra differ by 3dB!

More confusion arises when one paragraph of a textbook specifies S as a double sided intensity spectrum while in another paragraph it must be concluded³ from the thermal noise that S denotes a single sided spectrum.

Unambiguous definitions of S may sometimes be inferred based on physical quantities such as thermal noise⁴ or ideal shot noise⁵. With other authors the relationship to the auto-correlation⁶ function may give a clue. Many textbook discussions and publications use terms related to intensity spectra, that are employed in a rather loose fashion. It is the prime function of section 7.1 to provide the required definitions.

7.1.1. Spectral identification of random signals

Linear circuit analysis is relatively simple when using a spectral representation of the various signals. There is no unique definition of spectrum, since periodic and random signals have no Fourier transform due to the divergent Fourier integral when integrating from $-\infty$ to $+\infty$.

To ensure convergence of the Fourier integrals, signals must be switched on and off. This also holds when integrating the signal over a limited time interval, while assuming that the signal is zero outside this interval. Using this concept, the following spectra can be distinguished:

- For periodic signals, it is convenient to choose the length of the integration time interval equal to the period time of the signal. This yields the *discrete spectrum*, being the complex coefficients of the Fourier series expansion in complex form.
- For random signals, it is convenient to choose an interval T that is long enough to smooth the spectral rms-amplitude, averaged within a small frequency band $\Delta f > (1/T)$. This yields the *complex noise spectrum*.
- For a finite number of signal pulses, the total energy is finite. This means that integration from $-\infty$ to $+\infty$ yields convergent Fourier integrals and that the standard Fourier-transform is applicable. This provides the *continuous spectrum*.
- For auto-correlated random signals the total energy is also finite. There is an additional mathematical operation, to facilitate convergence of the Fourier integrals. This combined transformation yields the *intensity spectrum*, which is also called the power⁷ spectrum or the power spectral density.

³ An example is Dobrowolski [724]. Equation 5.4 of [724] relates the power spectral density $S(f)$ with the auto-correlation function, from which one concludes that it is a *double sided* intensity spectrum. From equation 5.7, 5.33 and 5.48 it is clear that the thermal noise current intensity of a resistor equals $S(f) = 4kT/R$. This intensity is equal to the *single sided* intensity spectrum. This illustrates the confusion related to power spectra.

⁴ Thermal noise is: $S_u(f) = 4kTR$ and $D_u(f) = 2kTR$ for voltages, or $S_i(f) = 4kT/R$ and $D_i(f) = 2kT/R$ for currents.

⁵ Shot noise current, associated with dc-current I_{dc} , is: $S_i(f) = 2q \cdot I_{dc}$ and $D_i(f) = q \cdot I_{dc}$

⁶ The auto-correlation function is: $R_u(t) = \langle U(t) \cdot U^*(t+\tau) \rangle_\infty$

⁷ We prefer the name *intensity* spectrum instead of *power* spectrum to avoid confusion with respect to the dimensions of the spectrum. Intensity spectra can be specified in A^2/Hz , in V^2/Hz or in W/Hz .

The mathematical definitions of all these spectra are summarized below. When there is no ground for confusion, the abbreviated notation will be used. Appendix K provides a compact summary of all related definitions and properties.

$$\begin{aligned}
 \mathbf{j}_d^* \{U(t); T\} &\stackrel{\text{def}}{=} \frac{1}{T} \cdot \int_{-1/2T}^{+1/2T} U(t) \cdot \exp(-j \cdot 2\pi n \cdot t/T) \cdot dt & Q_u(n) &= \text{discrete spectrum} \\
 \mathbf{j}_n \{U(t); T\} &\stackrel{\text{def}}{=} \frac{1}{\sqrt{1/2T}} \cdot \int_{-1/2T}^{+1/2T} U(t) \cdot \exp(-j \cdot 2\pi f \cdot t) \cdot dt & N_u(f) &= \text{complex noise spectrum} \\
 \mathbf{j}_c \{U(t)\} &\stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} U(t) \cdot \exp(-j \cdot 2\pi f \cdot t) \cdot dt & F_u(f) &= \text{continuous spectrum} \\
 \mathbf{D}\{U(t)\} &\stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} \langle U(t) \cdot U^*(t+\tau) \rangle_{\infty} \cdot \exp(-j \cdot 2\pi f \cdot \tau) \cdot d\tau & D_u(f) &= \text{double sided intensity spectrum} \\
 \mathbf{S}\{U(t)\} &\stackrel{\text{def}}{=} 2 \cdot \int_{-\infty}^{+\infty} \langle U(t) \cdot U^*(t+\tau) \rangle_{\infty} \cdot \exp(-j \cdot 2\pi f \cdot \tau) \cdot d\tau & S_u(f) &= \text{single sided intensity spectrum}
 \end{aligned}$$

In these formulae, $U^*(t)$ denotes the conjugate transpose⁸ of the signal $U(t)$. This operation is relevant when $U(t)$ is the (complex) mathematical transformed form of some physical (real) signal. In addition, $\langle f(t) \rangle_{\Delta t}$ denotes the signal average over interval Δt of function $f(t)$. See appendix K for an exact definition.

Note that the word *single* sided spectrum implies that negative frequencies are not used in the reverse transformation to the time domain, while the reverse transformation of the *double* sided spectrum uses positive and negative frequencies. It is a symmetrical function.

Noise spectra, averaged within a resolution bandwidth

The definition of complex noise spectra $N_u(f)$ enables simple relations between various signals and spectra. The most significant rules are embedded in the following well-known formulas:

$\langle N_u \cdot N_u^* \rangle_{\Delta f} \approx S_u(f) \quad \text{Wiener-Khinchine theorem}$
$\int_0^{\infty} S_u(f) \cdot df = \langle U(t) ^2 \rangle_{\infty} \quad \text{Parseval theorem}$

The Wiener-Khinchine theorem⁹ [703,704] holds for sufficiently small resolution bandwidth Δf , provided that the selected interval length T is sufficiently large ($T > 1/\Delta f$). For a proof of this theorem, see van der Ziel [710,720].

The Parseval theorem illustrates why S_u is called a *single sided* spectrum. The integral is restricted to positive frequencies. The same result would have been obtained with the *double sided* spectrum D_u when 'negative' frequencies are included for integration from $(-\infty \dots +\infty)$.

Spectrum analyzer measurements

Spectrum analyzers measure none of these spectra directly. These selective voltmeters, with top detector and resolution bandwidth Δf , indicate voltages similar to:

$$\text{when detecting harmonic}^{10} \text{ signals:} \quad \rightarrow \text{indicating:} \quad \approx |Q_u(f)|$$

⁸ Note that $U(t)$ is in this situation a scalar quantity, not a vector or matrix of scalar quantities. For scalar quantities holds that $U^*(t) \equiv U^*(t)$, in which $U^*(t)$ represents the complex conjugate of $U(t)$. For matrix quantities holds $U^*(t) \equiv U^{*T}(t)$. The use of a conjugated *transpose* in our definition simplifies a generalization of this definition.

⁹ We note that the original theorem states that $S_u \rightarrow 2 \cdot \int_{-\infty}^{+\infty} \langle U(t) \cdot U^*(t+\tau) \rangle_{\infty} \cdot \exp(-j \cdot 2\pi f \cdot \tau) \cdot d\tau$ and $S_u \stackrel{\text{def}}{=} \langle N_u \cdot N_u^* \rangle_{\Delta f}$

¹⁰ Harmonic signals originate from the composition of a finite number of periodical signals.

when detecting noisy signals: $\rightarrow$ indicating: $\approx |N_u(f)|\sqrt{\Delta f} \rightarrow \sqrt{S_u \cdot \Delta f}$

Modern spectrum analyzers, with built-in noise markers, automatically modify the readout to correct for the resolution bandwidth of the instrument. The voltage at the noise marker frequency approximates $|N_u(f)|$. The video filter averages the detected $|N_u(f)|$ over the resolution bandwidth.

The smaller the video bandwidth the more the noise marker indication approaches $\sqrt{S_u}$ when noisy signals are detected. The Wiener-Khintchine theorem predicts that replacement of the top-detector by a true-rms detector improves the accuracy of this result.

7.1.2. Spectral correlation between random signals

Consider noisy voltages $V_p(t)$, in which p denotes an index ranging from 1 to n . The associated complex noise spectra¹¹ are represented as $V_p(f)$. In most practical situations, *complex* noise spectra are not available from measurements. This is because the spectral magnitude and phase are rapidly fluctuating. More appropriate spectral quantities are intensity spectra since they represent averaged quantities.

According to the Wiener-Khintchine theorem the intensity spectrum is a good measure for the rms-magnitude of complex noise spectra, averaged in a small frequency band Δf . The *cross-correlation intensity spectrum* (cross-spectral intensity) is a similar quantity, and is a suitable measure for the average phase relation between two signals. When analyzing noisy circuits with many noise sources, their (cross) intensity spectra are most conveniently organized in matrix form. These spectra are defined¹² as:

$\mathbf{C} \stackrel{\text{def}}{=} \langle \mathbf{V} \cdot \mathbf{V}' \rangle_{\Delta f} = \begin{bmatrix} S_1 & S_{1 2} & \dots & S_{1 n} \\ S_{2 1} & S_2 & \dots & S_{2 n} \\ \dots & \dots & \dots & \dots \\ S_{n 1} & S_{n 2} & \dots & S_n \end{bmatrix}$	$S_{p q} \stackrel{\text{def}}{=} 2 \cdot \int_{-\infty}^{+\infty} \langle V_p(t) \cdot V_q'(t+\tau) \rangle_{\infty} \cdot \exp(-j \cdot 2\pi f \cdot \tau) \cdot d\tau$ $\gamma_{pq} \stackrel{\text{def}}{=} \gamma'_{qp} \stackrel{\text{def}}{=} S_{p q} / \sqrt{S_p \cdot S_q} = \frac{\langle V_p \cdot V_q' \rangle}{\sqrt{\langle V_p ^2 \rangle \cdot \langle V_q ^2 \rangle}}$ $C_{p q} = S_{p q} = \gamma_{pq} \cdot \sqrt{S_p} \cdot \sqrt{S_q} = \langle V_p(f) \cdot V_q'(f) \rangle_{\Delta f}$

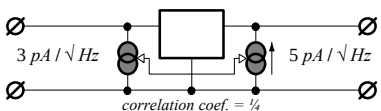
$\mathbf{V}(f)'$ is the conjugate transpose of $\mathbf{V}(f)$
 $\mathbf{C}$ = noise correlation matrix = $\mathbf{C}'$

$S_{p|q} = C_{p|q}$ = cross-correlation intensity spectrum
 $S_p = S_{p|p}$ = (self) intensity spectrum
 $\gamma_{p|q}$ = normalized correlation coefficient

The normalized correlation coefficients $\gamma_{p|q}$ are dimensionless, and the magnitudes of all these (complex) coefficients are bounded by $[0 \dots 1]$. These coefficients can also be organized in matrix format, termed as the *normalized noise correlation matrix* γ .

¹¹ In this case we use a shorthand notation, given by: $V(f) \stackrel{\text{def}}{=} N_V(f) \stackrel{\text{def}}{=} \int_n \{V(t); T\}$.

¹² The interpretation of these matrix definitions can be clarified using a model of a noisy amplifier. The model illustrated below shows two correlated noise sources, one at the amplifier input and the other at the output. The correlation matrix represents the spectral intensities of these sources in a well-organized way.



$$S_1 = (3 \text{ pA} / \sqrt{\text{Hz}})^2$$

$$S_2 = (5 \text{ pA} / \sqrt{\text{Hz}})^2$$

$$\gamma = 1/4$$

$$\mathbf{C} = \begin{bmatrix} 3^2 & 1/4 \cdot 3 \cdot 5 \\ 1/4 \cdot 3 \cdot 5 & 5^2 \end{bmatrix} [pA / \sqrt{Hz}]^2$$

The Wiener-Khintchine theorem in matrix notation $\mathbf{C} = \langle \mathbf{V} \cdot \mathbf{V}' \rangle_{\Delta f}$ provides the simple means to relate the measurable matrix quantity $\mathbf{C}$ to the theoretical vector quantity $\mathbf{V}$. This is an important property when applying circuit analysis to noisy circuits.

Other names for noise correlation matrix $\mathbf{C}$ that have been used from the beginning are: coherence matrix [Wiener: 709], noise spectral matrix [Haus and Adler: 716], power spectrum matrix [Gamo: 717], or the noise power matrix [Bosma: 719]. More recent publications tend to use the term *correlation matrix*, e.g. [Hillbrandt and Russer: 721].

7.1.3. Analytical noise analysis using complex noise spectra

Noise signals are usually quantified by spectral intensity and their correlation with other noisy signals. This is because rapid fluctuations of noisy signals interfere with reliable measurement of the phase of their complex noise spectra. It is, therefore, common practice to base noise analysis on intensity spectra.

It is unhandy when (manual) analyzing noisy circuits to use intensity spectra at each step of the analysis. But assuming that the magnitude and phase of complex noise spectra are *completely known*, all spectral analysis methods for deterministic signals can be applied to noisy signals. It is more convenient to ignore the lack of phase information, and act as if noisy signals are deterministic. Using this approach, the intensity spectrum of a noisy signal of interest is evaluated as follows:

- Identify each noise source with a unique symbol, such as $V(f)$, $J(f)$ or $W(f)$, representing the complex noise spectrum of that source.
- Derive an analytical expression for the output noise (or any other equivalent noise source of interest) using standard AC-analysis techniques.
- Transform this analytical expression for the complex noise spectrum into an expression for the intensity spectrum.

For instance, let $N(f) = \alpha \cdot V + \beta \cdot J$ be an expression, derived for the complex noise spectrum of interest. In this expression denote $V(f)$ and $J(f)$ the complex noise spectra of a voltage- and current noise source respectively. The circuit dependent variables $\alpha(f)$ and $\beta(f)$ are expressed in terms of resistance, capacitance, etc. The transformation of a complex noise spectrum $N(f)$ into its associated intensity spectrum $S(f)$ is:

$$\begin{aligned} S &= \langle N \cdot N' \rangle = \langle (\alpha \cdot V + \beta \cdot J) \cdot (\alpha' \cdot V' + \beta' \cdot J') \rangle \\ S &= \langle \alpha \cdot V \alpha' \cdot V' \rangle + \langle \beta \cdot J \beta' \cdot J' \rangle + \langle \alpha \cdot V \beta' \cdot J' \rangle + \langle \beta \cdot J \alpha' \cdot V' \rangle \\ S &= \alpha \cdot \alpha' \cdot \langle V \cdot V' \rangle + \beta \cdot \beta' \cdot \langle J \cdot J' \rangle + \alpha \cdot \beta' \cdot \langle V \cdot J' \rangle + \alpha' \cdot \beta \cdot \langle J \cdot V' \rangle \\ S &= \alpha \cdot \alpha' \cdot S_V + \beta \cdot \beta' \cdot S_J + \alpha \cdot \beta' \cdot S_{VJ} + \alpha' \cdot \beta \cdot S_{JV} \\ S &= |\alpha|^2 \cdot S_V + |\beta|^2 \cdot S_J + 2 \cdot \text{re} \{ \alpha \cdot \beta' \cdot S_{VJ} \} \end{aligned}$$

When numerical values for S_V , S_J , S_{VJ} and S_{JV} have been experimentally determined, the numerical value for S can be evaluated.

This analytical approach using complex noise spectra instead of intensity spectra is convenient for *manual* noise analysis of circuits. On the other hand, most circuit simulators provide numerical results instead of analytical expressions. The analytical approach is thus inconvenient for use with circuit simulators. Section 7.2.4 describes an alternative method using noise correlation matrices that is more suited to *automated* noise analysis.

7.2. Blackbox representation of noisy multi-port networks

The validation of equivalent circuit models is limited when modeling noisy linear devices with lumped noise sources. The more parasitic effects are taken into account, the wider the frequency interval will be in which the model is applicable.

Wideband analysis of noisy circuits requires a more robust approach. Blackbox representations of linear noisy circuits separate the physical origin of internal noise sources from what is externally perceptible. Blackbox methods are more widely applicable and provide a good starting-point for developing adequate device noise models.

Section 7.2 extends the blackbox analysis of linear circuits, as described in section 2.2, to include circuits with internal sources. It starts from the basic analysis of circuits comprising deterministic signal sources to construct a systematic analysis of arbitrary noisy circuits.

The voltage, current and wave methods described here are in essence known from the literature. In practice, publications focus on one or two of these three, confining the intercomparison of the methods. Furthermore, the wave methods discussed in the literature are often restricted to real reference impedances. One aim of this chapter is to discuss all the methods *simultaneously*, and to extend the application of the wave methods to ports normalized to arbitrary (complex) reference impedances.

7.2.1. Description of multi-port circuits including deterministic sources

Consider a multinode circuit for which all nodes are externally accessible via ports. A Norton representation of an internal signal source results in a current source. This current flow between two nodes, for instance directed from node n_1 to n_2 , is equivalent to re-routing the current flow via an arbitrary third node, e.g. the ground node. Therefore, a current source between two nodes is equivalent to two identical current sources: one directed from node n_1 to ground and another directed from ground to node n_2 . This is illustrated in figure 7.2.



Fig 7.2 A current source connected between nodes is equivalent to two identical current sources.

When the circuit comprises two or more sources, this approach holds for any individual source. Figure 7.3a shows the combined equivalent circuit model, being an extended Norton model. Each equivalent current source between a port node and ground originates from the summation of all current sources connected with that node and re-routed via ground.

This extended Norton model of multi-ports links up nicely when describing the multi-port with y-parameters. Figure 7.3a also shows the associated admittance matrix equation. Reorganization of these matrix equations results in the reverse situation, as is shown in figure 7.3b. It provides an extended Thévenin model, in which all current

sources are transformed into voltage sources. This model links up nicely with z-parameters. A similar approach holds for waves, as is shown in figure 7.3c.

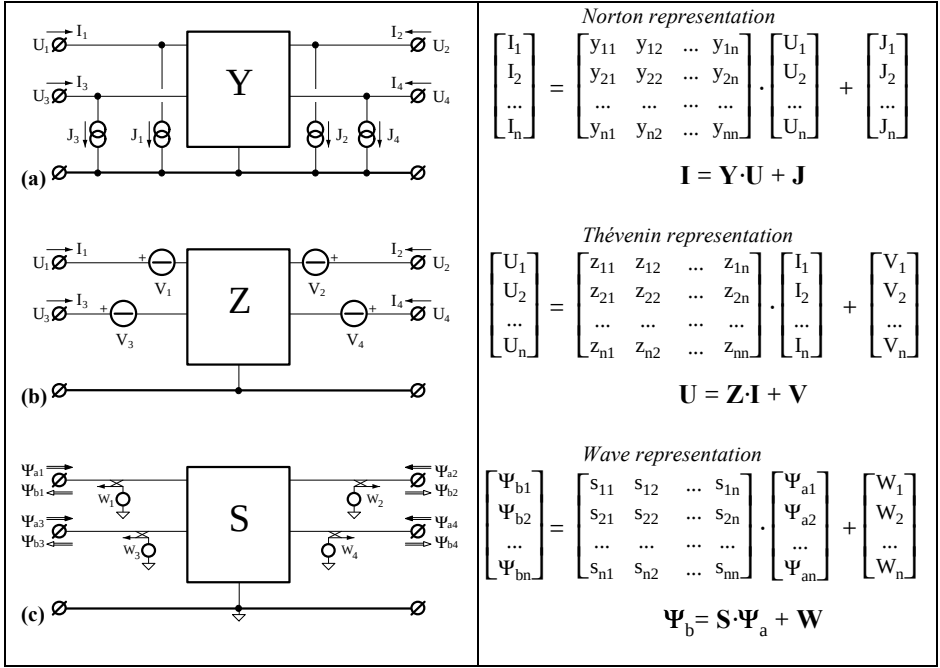


Fig 7.3 Three basic blackbox models of multi-port networks comprising arbitrary (deterministic) signal sources. They are a generalization of the Norton, Thévenin and wave representations of one-port sources, as described in section 2.1.3. The exchangeable source power is [715]:

$$P_e = \frac{1}{2} \cdot \mathbf{V}' / (\mathbf{Z} + \mathbf{Z}') \cdot \mathbf{V} = \frac{1}{2} \cdot \mathbf{J}' / (\mathbf{Y} + \mathbf{Y}') \cdot \mathbf{J} = \mathbf{W}' / (\mathbf{\sigma} - \mathbf{S} \cdot \mathbf{\sigma}' \cdot \mathbf{S}') \cdot \mathbf{W}$$

In section 2.2.1, transformation rules were summarized between y-, z- and s-parameters. The s-parameter transformations apply to waves normalized to arbitrary reference impedances Z_N (port normalization numbers). The reference impedances for the individual ports are applicable, including complex impedances with negative real part. Additional transformation rules for the equivalent sources $\mathbf{J}$, $\mathbf{V}$ and $\mathbf{W}$ are summarized in figure 7.4. They are simply reducible from the relations between voltages, currents and waves, as defined in section 2.1.1.

All transformation rules for sources have in common that the transformation reduces to a simple matrix product. This is an important property when extracting statistical properties of these (noise) sources.

	Y	Z	S
$\mathbf{J} = \Phi_{yz} \cdot \mathbf{V}$	$\Phi_{yz} = -\mathbf{Y} = -\text{inv}(\mathbf{Z}) = \mathbf{r}_N \cdot ((\mathbf{S} \cdot \mathbf{Z}_N + \mathbf{Z}_N') \cdot (\mathbf{1} - \mathbf{S})) / \mathbf{r}_N$		
$\mathbf{V} = \Phi_{zy} \cdot \mathbf{J}$	$\Phi_{zy} = -\text{inv}(\mathbf{Y}) = -\mathbf{Z} = \mathbf{r}_N \cdot ((\mathbf{1} - \mathbf{S})(\mathbf{S} \cdot \mathbf{Z}_N + \mathbf{Z}_N')) / \mathbf{r}_N$		
$\mathbf{J} = \Phi_{ys} \cdot \mathbf{W}$	$\Phi_{ys} = -(\mathbf{Y} \cdot \mathbf{Z}_N + \mathbf{1}) / (\sigma \cdot \mathbf{r}_N) = -(\mathbf{Z} \cdot \mathbf{Z}_N + \mathbf{1}) / (\sigma \cdot \mathbf{r}_N) = -(2 \cdot \mathbf{r}_N) / (\mathbf{S} \cdot \mathbf{Z}_N + \mathbf{Z}_N')$		
$\mathbf{W} = \Phi_{sy} \cdot \mathbf{J}$	$\Phi_{sy} = -(\sigma \cdot \mathbf{r}_N) / (\mathbf{Y} \cdot \mathbf{Z}_N + \mathbf{1}) = -(\sigma \cdot \mathbf{r}_N) / (\mathbf{Z} \cdot \mathbf{Z}_N + \mathbf{1}) = -(\mathbf{S} \cdot \mathbf{Z}_N + \mathbf{Z}_N') / (2 \cdot \mathbf{r}_N)$		
$\mathbf{V} = \Phi_{zs} \cdot \mathbf{W}$	$\Phi_{zs} = (\mathbf{Z}_N + \mathbf{Y} \cdot \mathbf{1}) / (\sigma \cdot \mathbf{r}_N) = (\mathbf{Z}_N + \mathbf{Z}) / (\sigma \cdot \mathbf{r}_N) = (2 \cdot \mathbf{r}_N) \cdot (\mathbf{1} - \mathbf{S})$		
$\mathbf{W} = \Phi_{sz} \cdot \mathbf{V}$	$\Phi_{sz} = (\sigma \cdot \mathbf{r}_N) / (\mathbf{Z}_N + \mathbf{Y} \cdot \mathbf{1}) = (\sigma \cdot \mathbf{r}_N) / (\mathbf{Z}_N + \mathbf{Z}) = (\mathbf{1} - \mathbf{S}) / (2 \cdot \mathbf{r}_N)$		

Fig 7.4 Transformation rules for the equivalent sources of figure 7.3. All of these transformations share the form: $\mathbf{X}_b = \Phi_{ba} \cdot \mathbf{X}_a$. For the definition of σ , and $\mathbf{r}_N$ see section 2.2.1. The transformation matrices using s-parameters hold for waves normalized to arbitrary reference impedances Z_n .

7.2.2. Description of noisy multi-ports using correlation matrices

There are many blackbox descriptions for describing noisy multi-ports. More convenient methods have in common that noise and transfer descriptions are closely matched. When matrix descriptions are chosen, standard blackbox models are preferred, similar to figure 7.3. This requires the replacement of deterministic sources by noise sources, and the representation of these equivalent sources by complex noise spectra $\mathbf{J}$, $\mathbf{V}$ or $\mathbf{W}$.

Complex noise spectra are primarily intended for use as intermediate results and are not suitable for quantify noise sources. Self- and cross-correlation intensity spectra are more appropriate, and can be organized in correlation matrix format. For Y-, Z- and S-matrix transfer descriptions we define¹³ the associated noise correlation matrices as follows:

$\mathbf{C}_Y \stackrel{\text{def}}{=} \langle \mathbf{J} \cdot \mathbf{J}' \rangle_{\Delta f}$	= noise current correlation matrix
$\mathbf{C}_Z \stackrel{\text{def}}{=} \langle \mathbf{V} \cdot \mathbf{V}' \rangle_{\Delta f}$	= noise voltage correlation matrix
$\mathbf{C}_S \stackrel{\text{def}}{=} \langle \mathbf{W} \cdot \mathbf{W}' \rangle_{\Delta f}$	= noise wave correlation matrix

Let's consider an arbitrary circuit topology, with various uncorrelated noise sources. Common circuit theory is applicable when the individual noise sources are represented by complex noise spectra. Then, perform the same calculation as with deterministic sources. This to extract the desired transfer matrix of the blackbox representation, and its associated equivalent noise sources. Finally, convert¹⁴ their complex noise spectra into the desired correlation matrix format. This approach is much less complicated for manual analysis than using correlated intensity spectra at each stage of the analysis.

¹³ Some authors use definitions that are 'normalized' with respect to thermal noise: all matrices are divided by kT.

¹⁴ For instance, let J_1 and J_2 be equivalent noise sources of a Norton representation of a circuit. They are linearly related to some uncorrelated internal sources J_a and J_b with intensity spectra S_a and S_b . The element value $\langle J_1 \cdot J_2' \rangle$ of the admittance noise correlation matrix C_Y is extracted according to:

$$\left. \begin{aligned} J_1 &= \alpha_1 \cdot J_a + \alpha_2 \cdot J_b \\ J_2 &= \alpha_3 \cdot J_a + \alpha_4 \cdot J_b \end{aligned} \right\} \Rightarrow \langle J_1 \cdot J_2' \rangle = \alpha_1 \cdot \alpha_3' \cdot S_a + \alpha_2 \cdot \alpha_4' \cdot S_b + 0 + 0_a$$

Using noise correlation matrices associated with y-, z- or s-parameters has many advantages. These advantages are decisive to simplify automated noise analysis on arbitrary circuits.

At first, the composition of C_Y for arbitrary circuits can be performed in a plain way. In addition, it is easy to convert C_Y into C_Z or C_S using simple matrix products. Furthermore, when most nodes of a circuit are not externally available as port nodes, the reduction of the correlation matrix dimension is performed similar to that of transfer matrices. The are elucidated below.

Composition rule for C_Y

Matrix C_Y of a complex noisy circuit originates from the matrix addition of correlation matrices of all its individual devices. This is because (1) shunted current sources will add and (2) the noise of one device is uncorrelated with the noise of another one. As a result:

$$C_{Y,\text{circuit}} = \sum c_{Y,\text{devices}}$$

This composition rule is restricted to admittance noise correlation matrices only. Its application in circuit simulators will be discussed in section 7.2.4.

Transformation rules for correlation matrices

The transformation rules of figure 7.4 for equivalent currents to equivalent voltages or waves are also applicable to complex noise spectra. Let $\mathbf{X}_a$ be a vector representing equivalent sources to be transformed into vector $\mathbf{X}_b$. The transformation of an arbitrary $\mathbf{X}_a$ into $\mathbf{X}_b$ equals: $\mathbf{X}_b = \Phi_{ba} \cdot \mathbf{X}_a$, in which matrix Φ is the associated transformation matrix.

This formula provides the transformation rule of an arbitrary noise correlation matrix C_a to C_b and has the following general form:

$$C_b = \langle \mathbf{X}_b \cdot \mathbf{X}_b' \rangle = \langle (\Phi_{ba} \cdot \mathbf{X}_b) \cdot (\Phi_{ba} \cdot \mathbf{X}_b)' \rangle = \Phi_{ba} \cdot \langle \mathbf{X}_b \cdot \mathbf{X}_b' \rangle \cdot \Phi_{ba}'$$

$$C_b = \Phi_{ba} \cdot C_a \cdot \Phi_{ba}'$$

The general transformation matrices Φ in figure 7.4 are all applicable to transforming noise correlation matrices. The most important relations are:

$$\begin{aligned} C_Z &= Z \cdot C_Y \cdot Z' \\ C_Y &= Y \cdot C_Z \cdot Y' \\ C_S &= Q \cdot C_Z \cdot Q' \quad Q = (1-S)/(2 \cdot r_N) \end{aligned}$$

Correlation matrices have been known for years. In addition, Haus and Adler [715,716] derived in 1958 the natural regularity of the above transformation, as intermediate result in discussing properties for noise measure. Bosma [719] derived in 1967 the same regularity, when treating these transformations as mathematical coordinate transformations in a signal-state space. Just in 1976, Hillbrand and Russer [721] illustrated the importance of correlation matrices for the analysis of noisy circuits. They showed how to apply these matrices for simplifying two-port noise analysis.

7.2.3. Reduction of correlation matrix dimension of noisy multi-ports

Consider a circuit with N_n nodes, represented by an $N_n \times N_n$ matrix (y-, z- or s-parameters). In many circuit applications, no more than a few nodes ($N_p < N_n$) are externally available, for instance nodes p_1 and p_2 (the port nodes). This means that all Thévenin voltage sources of figure 7.3b are floating on one side, except those connected to the port nodes. As a result, these voltage sources are unused and can be omitted. In section 2.2.3 a similar situation was discussed. It explained that the matrix reduction of the *Thévenin* representation is simply the deletion of all rows and columns associated with internal nodes.

Let $\mathbf{U} = \mathbf{Z} \cdot \mathbf{I} + \mathbf{V}$ be the original matrix representation of a multi-port network, and let $\mathbf{u} = \mathbf{z} \cdot \mathbf{i} + \mathbf{v}$ be its reduced matrix representation. The reduced signal vectors $\mathbf{u}$, $\mathbf{i}$ and $\mathbf{v}$ originate from $\mathbf{U}$, $\mathbf{I}$ and $\mathbf{V}$ by deletion of all rows, except those with number p_1 and p_2 . The same applies for the rows and columns of $\mathbf{Z}$ to generate $\mathbf{z}$. Using the port definition matrix $\mathbf{P}$ (see section 2.2.3) for simplifying this reduction formalism yields:

$$\mathbf{P} \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{array}{l} \leftarrow \text{row } p1 \\ - \\ \leftarrow \text{row } p2 \\ - \\ - \\ - \end{array} \quad \Rightarrow \quad \begin{cases} \mathbf{i} = \mathbf{P}' \cdot \mathbf{I} \\ \mathbf{u} = \mathbf{P}' \cdot \mathbf{U} \\ \mathbf{v} = \mathbf{P}' \cdot \mathbf{V} \\ \mathbf{z} = \mathbf{P}' \cdot \mathbf{Z} \cdot \mathbf{P} \end{cases}$$

This matrix identity is reserved for the Thévenin representation only (z-parameters). Reduction of y- and s-parameters requires z-parameters as intermediate representation format. Use the transformation rules of figure 7.4 for this purpose, for instance: $\mathbf{y} = \text{inv}(\mathbf{P}' \cdot \mathbf{Y} \cdot \mathbf{P})$ and $\mathbf{i} = (\mathbf{y} \cdot \mathbf{P}') \cdot (\mathbf{Y} \cdot \mathbf{I})$.

A similar reduction technique can be performed with correlation matrices. Consider a circuit with N_n nodes, of which the voltage noise correlation matrix is known. Let $\mathbf{C}_z$ be a correlation matrix and $\mathbf{V}$ be a noisy voltage source vector to be reduced to matrix $\mathbf{c}_z$ and vector $\mathbf{v}$. Using the above port definition matrix $\mathbf{P}$, the reduction of the source vector is performed with the multiplication $\mathbf{v} = \mathbf{P}' \cdot \mathbf{V}$. The reduced correlation matrix $\mathbf{c}_z$ is:

$$\mathbf{c}_z = \langle \mathbf{v} \cdot \mathbf{v}' \rangle = \langle (\mathbf{P}' \cdot \mathbf{V}) \cdot (\mathbf{P}' \cdot \mathbf{V})' \rangle = \mathbf{P}' \cdot \langle \mathbf{V} \cdot \mathbf{V}' \rangle \cdot \mathbf{P}$$

$$\mathbf{c}_z = \mathbf{P}' \cdot \mathbf{C}_z \cdot \mathbf{P}$$

This reduction rule is restricted to voltage noise correlation matrices only. Reduction of $\mathbf{C}_y$ and $\mathbf{C}_s$ requires $\mathbf{C}_z$ -parameters as intermediate representation format. Use the transformation matrices of figure 7.4 for this purpose.

All these reduction techniques are important when implementing circuit simulators as described in section 2.2.4 and section 7.2.4.

7.2.4. Application of noise correlation matrices in circuit simulators.

Circuit simulators can easily be extended with noise analysis facilities when using y-parameters for transfer analysis and $\mathbf{C}_y$ correlation matrices for noise analysis. This method has been demonstrated in 1985 by Rizzoli and Lipparini [723]. Because our integrated approach of designing and characterizing wideband circuits relies extensively on various computer algorithms, this section discusses algorithms for automated analysis of noisy circuits in more detail.

Consider an arbitrary circuit topology, of which all elements are subsequently described in a netlist. The admittance matrix $\mathbf{Y}$ of this circuit is constructed as discussed in section 2.2.4. The associated correlation matrix $\mathbf{C}_y$ is constructed simultaneously by insertion of each individual noise source, as described below:

- An uncorrelated current noise source from node n_1 to n_2 with spectral intensity S_i is modeled by two fully correlated sources. One current flowing from node n_1 to ground, and another flowing with opposite sign from node n_2 to ground. This has been illustrated in figure 7.2. The associated device correlation matrix $\mathbf{C}_y$ is as follows added to the circuit correlation matrix $\mathbf{C}_y$:

$$\begin{bmatrix} c_{n2,n1} & c_{n2,n3} \\ c_{n4,n1} & c_{n4,n3} \end{bmatrix} = \begin{bmatrix} c_{n2,n1} & c_{n2,n3} \\ c_{n4,n1} & c_{n4,n3} \end{bmatrix} + S_i \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

- Two or more correlated current noise sources are inserted similarly to a single uncorrelated source. The difference is that the device correlation matrix $\mathbf{c}_y$ is a 4×4 matrix. It represents the (cross) intensity spectra of four equivalent sources from the four nodes to ground.

Let S_a , S_b , S_{ab} and S_{ba} be the associated (cross) intensity spectra ($S_{ba} = S'_{ab}$). A simple way to expand the 4×4 device correlation matrix is provided by the matrix product:

$$\mathbf{c}_y = \begin{bmatrix} +1 & -1 & 0 & 0 \\ 0 & 0 & +1 & -1 \end{bmatrix}' \cdot \begin{bmatrix} S_a & S_{ab} \\ S_{ba} & S_b \end{bmatrix} \cdot \begin{bmatrix} +1 & -1 & 0 & 0 \\ 0 & 0 & +1 & -1 \end{bmatrix} \quad (\mathbf{Z}' \text{ is the transpose of } \mathbf{Z})$$

- An uncorrelated voltage noise source, in series with a non-zero impedance, is inserted in $\mathbf{C}_y$ by using its Norton current representation. When this series impedance is zero, a non-zero impedance is imitated with two resistors: $+R$ and $-R$. The positive resistor is subsequently used for converting voltage sources into current sources. The consequence of this approach is allocating additional (internal) nodes.
- Thermal noise of resistors R can be inserted like uncorrelated noise currents. Each resistor is shunted with a noise current with intensity spectrum $S_i = 4kT/R$. The same applies for shot noise sources with intensity $S_i = 2q \cdot I_{dc}$ due to dc-currents I_{dc} . Insertion of thermal noise of a passive reciprocal sub-circuit in thermal equilibrium, such as a lossy transmission line, can be performed by evaluating $\mathbf{c}_y$ of the sub-circuit and addition of $\mathbf{c}_y$ to $\mathbf{C}_y$. Section 7.2.5 discusses how to evaluate $\mathbf{c}_y$ from the transfer matrix $\mathbf{Y}$.

When all noise sources of the net-list are inserted, the combined correlation matrix $\mathbf{C}_y$ is complete. The matrix $\mathbf{C}_z = \mathbf{Z} \cdot \mathbf{C}_y \cdot \mathbf{Z}'$ provides the associated Thévenin representation of the noisy circuit. The circuit simulator must now reduce this matrix into a smaller correlation matrix, for instance a 2×2 matrix $\mathbf{c}_z$ of a two-port. As described in section 7.2.3 the reduced correlation matrix $\mathbf{c}_z$ equals:

$$\mathbf{c}_z = \mathbf{P}' \cdot \mathbf{C}_z \cdot \mathbf{P} = \mathbf{P}' \cdot (\mathbf{Z} \cdot \mathbf{C}_y \cdot \mathbf{Z}') \cdot \mathbf{P}$$

$$\mathbf{c}_z = (\mathbf{P}' \cdot \mathbf{Z}) \cdot \mathbf{C}_y \cdot (\mathbf{P}' \cdot \mathbf{Z})'$$

The circuit simulator has completed when all $\mathbf{c}_z$ matrices are evaluated for all frequencies of interest. When $\mathbf{c}_y$ - or $\mathbf{c}_s$ -matrices are preferred, convert all $\mathbf{c}_z$ matrices into the preferred format. The dimension of matrix $\mathbf{c}_z$ depends on the number of ports of the circuit, and may differ from two-ports.

We extended the circuit simulator of section 2.2.4 with noise analysis facilities in the 4GL computer language MatLab® [123], and found that less then one or two additional pages source code are adequate for implementing the overall simulator.

7.2.5. Generalized thermal noise theorem for multi-port networks

In a practical situation, there is no need for measuring the thermal noise of a resistor. A simple resistance and temperature measurement is adequate for predicting this noise using the well-known Nyquist theorem [702]: $S_u = 4 \cdot kT \cdot R$. Complete proofs of this theorem can also be found in van der Ziel [710,720].

A generalized theorem facilitates the thermal noise calculation for complex impedances Z . The thermal noise voltage can be calculated from impedance measurements using $S_u = 4 \cdot kT \cdot \text{re}\{Z\}$.

This simple thermal noise theorem can further be generalized for arbitrary passive multi-ports in thermal equilibrium, e.g. two-port filters with lossy inductors and distributed (lossy) elements. It facilitates the evaluation of all three basic equivalent noise models shown in figure 7.3, merely from transfer measurements. Noise measurements are not required.

The thermal noise correlation matrices, associated with the equivalent noise models of figure 7.3, of passive reciprocal multi-ports in thermal-equilibrium are:

$\begin{aligned} \mathbf{C}_y &= \langle \mathbf{J} \cdot \mathbf{J}' \rangle_{\Delta f} &= 2kT \cdot (\mathbf{Y} + \mathbf{Y}') \\ \mathbf{C}_z &= \langle \mathbf{V} \cdot \mathbf{V}' \rangle_{\Delta f} &= 2kT \cdot (\mathbf{Z} + \mathbf{Z}') \\ \mathbf{C}_s &= \langle \mathbf{W} \cdot \mathbf{W}' \rangle_{\Delta f} &= kT \cdot (\boldsymbol{\sigma} - \mathbf{S} \cdot \boldsymbol{\sigma} \cdot \mathbf{S}') \end{aligned}$	$\begin{aligned} \rightarrow S_j &= 4kT \cdot \text{re}\{Y\} \\ \rightarrow S_v &= 4kT \cdot \text{re}\{Z\} \\ \rightarrow S_w &= kT(1 - \Gamma ^2) \end{aligned}$

k = Boltzmann constant, $1.380662 \cdot 10^{-23} [\text{J/K}]$

T = absolute temperature, in $[\text{K}]$

for definition of $\boldsymbol{\sigma}$ see section 2.2.1

spectral intensity of
one-port device

Proof

The calculation methods of section 7.2.4 for constructing correlation matrices are the bases for proofing the above generalized thermal noise theorems.

When constructing the admittance transfer matrix $\mathbf{Y}$ and the associated current correlation matrix $\mathbf{C}_y$ of arbitrary passive noisy circuits, one can apply the methods that are discussed in section 2.2.4 and 7.2.4 for circuit simulators. The relation between $\mathbf{C}_y$ and $\mathbf{Y}$ for arbitrary passive circuits is derived in the following steps:

- (a) The thermal noise contribution from individual resistors R equals: $S_i = 4 \cdot kT/R$. The way their admittance ($1/R$) is inserted in matrix $\mathbf{Y}$ equals the way their noise S_i is inserted in matrix $\mathbf{C}_y$. As a result, $\mathbf{C}_y = 4 \cdot kT \cdot \mathbf{Y}$ when all circuit elements are resistors at equal temperature.

- (b) Inductors and capacitors do not contribute to the generation of thermal noise, and their admittance is pure imaginary. As a result, $C_y = 4 \cdot kT \operatorname{re}\{\mathbf{Y}\}$ when all circuit elements are resistors, capacitors and inductors at equal temperature.
- (c) These passive networks are all reciprocal networks, see section 2.2.2, which means that $\mathbf{Y}' = \operatorname{conj}\{\mathbf{Y}^T\} = \operatorname{conj}\{\mathbf{Y}\}$ and that $\operatorname{re}\{\mathbf{Y}\} = (\mathbf{Y} + \mathbf{Y}')$. As a result, $C_y = 2 \cdot kT \cdot (\mathbf{Y} + \mathbf{Y}')$ for passive circuits in thermal equilibrium.

Using the transformation rules for the matrices $\{\mathbf{Y}, C_y\}$ to $\{\mathbf{Z}, C_z\}$ or $\{\mathbf{S}, C_s\}$, as discussed in section 2.2.1 and 7.2.2, provides the associated correlation matrices.

Validity

These expressions also hold for networks of which no more than a few nodes are externally available (port nodes). This can easily be verified for C_z since the reduction formula for matrix C_z (see section 7.2.3) equals the reduction formula for matrix $\mathbf{Z}$ (see section 2.2.3).

To the author knowledge, all publications evaluating C_s for thermal noise are restricted to waves normalized to complex reference impedances with *positive* real part. This study resulted in a generalized expression for matrix C_s that holds for any set of reference impedances, including complex values with *negative* real part.

For very high frequencies and low temperatures, an extension to the Nyquist theorem is required. All above thermal noise relations must be multiplied [710,720] with the Planck factor $p(f,T)$ being:

$$p(f,T) = \frac{(hf/kT)}{\exp(hf/kT)-1} \quad h = 6.626176 \cdot 10^{-34} \text{ [J.s]} = \text{Planck constant}$$

At room temperature this correction becomes significant for frequencies higher than 6000 GHz. Therefore, this correction may be omitted for many applications.

History

Thermal noise relations for C_y and C_z are known since 1955 by Twiss [711], as generalization of the Nyquist theorem [702]. Haus and Adler [716] derived in 1959 the same results. The eigenvalues of this matrix are invariant to lossless transformations that preserve the number of ports.

The relation for C_s for passive two-ports has been discussed in English¹⁵ as early as 1959 by Gamo [717]. Several years later, Bosma [719] discussed similar topics¹⁶.

An alternative way of proving the thermal noise relations for C_s is discussed in [725]

7.2.6. Conclusions

Various matrix methods have been summarized to represent noisy linear networks as blackbox. The review was focused on a simultaneous discussion of using z -, y - and s -parameters, without any restrictions on the reference impedance Z_N . It resulted in generalized Norton- Thévenin- and wave representations of (noise) sources.

¹⁵ Gamo refers to a Japanese publication from Takahasi (1952) on a generalization of the Nyquist theorem. This might be the first publication on this topic.

¹⁶ We observed that many authors referred to Bosma as the first who derived an expression for C_s . Some of them identify this expression as Bosma's theorem. Historically, there are no grounds for this designation.

The use of correlation matrices has many advantages. They facilitate transformation between various representation formats with ease, their dimension is simple to reduce, they facilitate a generalized description of thermal noise and they facilitate the simulation of arbitrary noisy circuits. The simplicity of implementing circuit simulator dealing with noise has been discussed.

Although these methods are essentially known from the literature, many articles and textbook discussions have contributed to produce this detailed overview. This is because many textbook discussions are focused on isolated topics or discuss methods that are restricted to real reference impedances or to positive complex reference impedances. Note that some authors use different definitions.

7.3. Two-port noise parameters

An important group of multi-port networks are two-ports. Automated measurement setups, dedicated to full two-port noise measurements at microwave frequencies, have come to the market. Moreover, transistor manufacturers are (sparingly) providing measured two-port noise parameters of their devices, to complement measured transfer parameters. These developments enable the application of two-port noise parameters as integral part of a design strategy for wideband amplifiers.

One can distinct different groups of noise parameters, as is summarized below:

- Designers of low-noise microwave circuits are using noise parameters for years. Their design goal is minimizing the noise factor, as explained in section 7.3.2, which can successfully be realized within small¹⁷ frequency bands. *Spot noise parameters* are dedicated to these narrow-band design goals, and therefore commonly used.
- Spot noise parameters are inconvenient when designing wideband amplifiers. *Matrix noise parameters* are more appropriated and have the additional advantage that the underlying concepts are also applicable to multi-port circuits.
- Designers of wideband amplifiers preferably use equivalent circuit models, extracted from noise measurements. This study resulted in the development of *autonomous noise parameters* that simplify the extraction of adequate noise models for synthesis purposes.

This section 7.3 discusses all these noise parameters to facilitate input noise calculations, parameter conversion and parameter measurement. The first two groups are essentially known concepts, although this study generalized the validation of the spot noise parameters. Autonomous noise parameters are promising new concepts.

7.3.1. Matrix noise parameters, dedicated to two-ports

All previously defined Norton-, Thévenin- and wave representations in matrix format are applicable to representing two-port noisy networks. Additionally, many other matrix representations are in use.

Using multiple noise representations can sometimes lead to more efficient analysis, as has been demonstrated by Hillbrandt and Russer [721]. Additionally, Hartmann [722] distinct twelve matrix representations dedicated to two-port networks. Figure 7.5 shows various representations of two-ports with two equivalent sources, each associated with a preferred matrix format.

¹⁷ The word 'small' is related to the center frequency of the frequency band of interest. Ten percent of the center frequency f_c is relatively small, although this is more than 1 GHz bandwidth at $f_c=10\text{GHz}$ center frequency. The word 'spot' is used to emphasize that the application of spot noise parameters is primarily focussed on single frequencies.

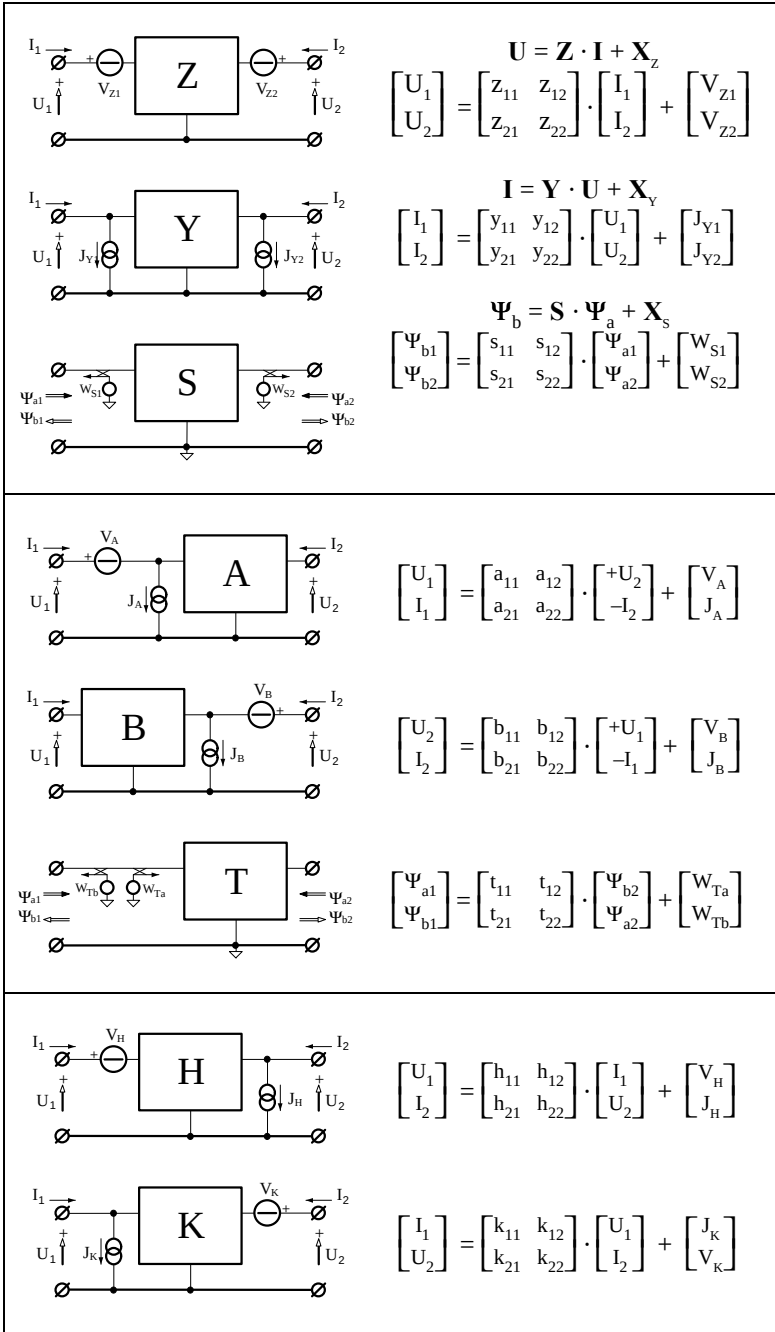


Fig 7.5 Rerepresentations of internal (noise) sources of two-ports with two equivalent sources. The overall calculation effort is simplified when the associated matrix representation is used.

Equivalent input noise correlation matrices

The most important two-port noise representations are those that have all equivalent (noise) sources transformed to the input port. They are preferably associated with chain parameters (A-matrix) or transmission parameters (T-matrix) to harmonize noise and transfer calculations. Their definition is:

$$\mathbf{C}_A \stackrel{\text{def}}{=} \begin{bmatrix} \langle \mathbf{V}_A \cdot \mathbf{V}'_A \rangle & \langle \mathbf{V}_A \cdot \mathbf{J}'_A \rangle \\ \langle \mathbf{J}_A \cdot \mathbf{V}'_A \rangle & \langle \mathbf{J}_A \cdot \mathbf{J}'_A \rangle \end{bmatrix} = \begin{bmatrix} S_v & S_{v|j} \\ S_{j|v} & S_j \end{bmatrix} \quad \text{for voltages and currents}$$

$$\mathbf{C}_T \stackrel{\text{def}}{=} \begin{bmatrix} \langle \mathbf{W}_A \cdot \mathbf{W}'_A \rangle & \langle \mathbf{W}_A \cdot \mathbf{W}'_B \rangle \\ \langle \mathbf{W}_B \cdot \mathbf{W}'_A \rangle & \langle \mathbf{W}_B \cdot \mathbf{W}'_B \rangle \end{bmatrix} = \begin{bmatrix} S_a & S_{a|b} \\ S_{b|a} & S_b \end{bmatrix} \quad \text{for waves}$$

These input noise parameters are measurable, independently from measurements of two-port transfer parameters. Specifying device noise as $\mathbf{C}_A$ or $\mathbf{C}_T$, enables the analysis of optimal noise performance without any knowledge on device transfer. $\mathbf{C}_A$ and $\mathbf{C}_T$ are input noise parameters among many others, see section 7.3.2. Nevertheless, these parameters yield in the most regular expressions for analyzing equivalent input noise.

Two-port transformation rules for correlation matrices

In section 7.2.1 various transformation rules are discussed for equivalent sources in multi-port networks. These transformation rules have $(\mathbf{X}_B = \Phi_{BA} \cdot \mathbf{X}_A)$ as general form. This regular form is maintained when the matrix representations of figure 7.5 are used. The associated two-port transformation matrices Φ are summarized in figure 7.6.

$\mathbf{X}_B = \Phi_{BA} \cdot \mathbf{X}_A$	$\mathbf{X}_Y$	$\mathbf{X}_Z$	$\mathbf{X}_A$	$\mathbf{X}_S$	$\mathbf{X}_T$
$\Phi_{Y\otimes} =$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -y_{11} & -y_{12} \\ -y_{21} & -y_{22} \end{bmatrix}$	$\begin{bmatrix} -y_{11} & 1 \\ -y_{21} & 0 \end{bmatrix}$	$(2 \cdot \mathbf{r}_N \cdot \mathbf{Y}) / (1 - \mathbf{S})$	
$\Phi_{Z\otimes} =$	$\begin{bmatrix} -z_{11} & -z_{12} \\ -z_{21} & -z_{22} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -z_{11} \\ 0 & -z_{21} \end{bmatrix}$	$(2 \cdot \mathbf{r}_N) / (1 - \mathbf{S})$	
$\Phi_{A\otimes} =$	$\begin{bmatrix} 0 & a_{12} \\ 1 & a_{22} \end{bmatrix}$	$\begin{bmatrix} 1 & -a_{11} \\ 0 & -a_{21} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$		
$\Phi_{S\otimes} =$	$(1 - \mathbf{S}) / (2 \cdot \mathbf{r}_N \cdot \mathbf{Y})$	$(1 - \mathbf{S}) / (2 \cdot \mathbf{r}_N)$		$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -s_{11} & 1 \\ -s_{21} & 0 \end{bmatrix}$
$\Phi_{T\otimes} =$				$\begin{bmatrix} 0 & -t_{11} \\ 1 & -t_{21} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Fig 7.6 Matrices to transform the sources of figure 7.5 from format $\mathbf{X}_A$ to format $\mathbf{X}_B$. For the definition of $\mathbf{r}_N$, see section 2.2.1. The transformation matrices using s-parameters hold for waves normalized to arbitrary reference impedances Z_n

Hillbrandt and Russer [721] discussed this regularity for Y-, Z- and A-matrices to demonstrate how matrix noise parameters simplify two-port noise analysis. When cascading two-ports, specified in A- or T- matrices, the input noise correlation matrices become (assuming equal reference impedance Z_0):

$$\begin{aligned} \mathbf{C}_A &= \mathbf{C}_{A1} + \mathbf{A}_1 \cdot \mathbf{C}_{A2} \cdot \mathbf{A}'_1 \\ \mathbf{C}_T &= \mathbf{C}_{T1} + \mathbf{T}_1 \cdot \mathbf{C}_{T2} \cdot \mathbf{T}'_1 \end{aligned}$$

Equivalent input excess noise of two-ports

Figure 7.7a and 7.7b show two equivalent noise models representing the noise of two-ports observed from their input port. These are (partly) correlated sources.

To simplify manual circuit analysis using partly correlated noise sources, it is convenient to replace them by sources that are fully correlated and by other sources that are fully uncorrelated. Figure 7.7c, -d and -e illustrate this concept in various representations. The correlation factors Y_c , Z_c and Γ_c represent the complex ratio between two fully correlated sources, as illustrated in the associated correlation matrices C_A and C_T .

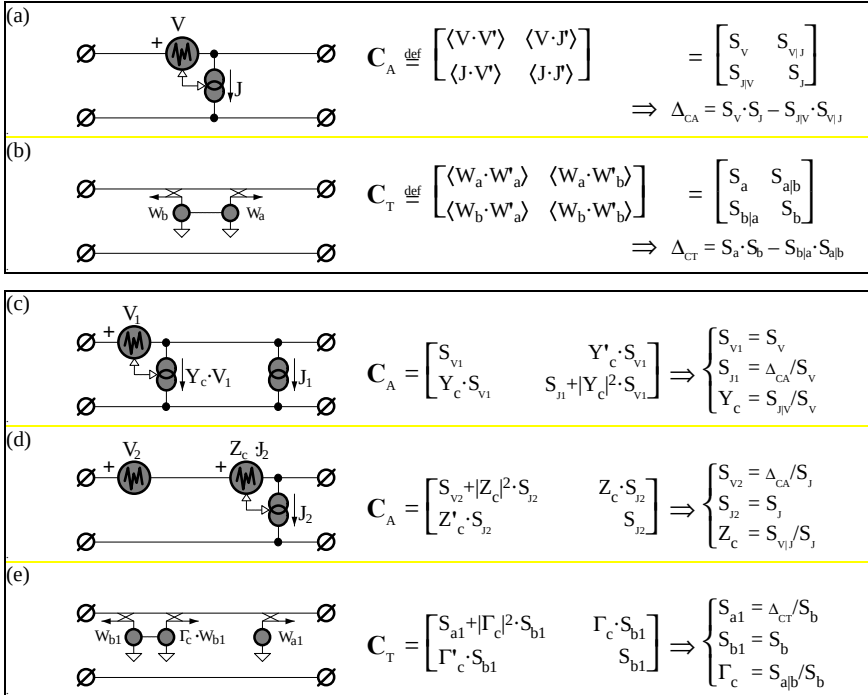


Fig 7.7 Two (partly) correlated noise sources can be represented by two fully correlated sources and an additional uncorrelated noise source. This approach simplifies manual circuit analyses of input noise.

An important application of the models in figure 7.7 is the calculation of the input excess-noise level when the noisy two-port input is connected with a source impedance. Excess noise is the device noise in *excess* to the thermal noise of the source impedances. It equals the total input noise when the source impedance is cooled down to temperature $T=0$.

Figure 7.8 quantifies the input excess noise as current, voltage or wave, in various ways. This excess noise is the combination of all equivalent device noise sources transformed to the input port. When the total input noise is requested, add the thermal noise of the source impedance. This is one of:

$$\begin{aligned} S_{is} &= 4kT \cdot \text{re}\{Y_s\} && \text{thermal noise current} \\ S_{us} &= 4kT \cdot \text{re}\{Z_s\} && \text{thermal noise voltage} \\ S_{ws} &= kT \cdot (1 - |\Gamma_s|^2) && \text{thermal noise wave} \end{aligned}$$

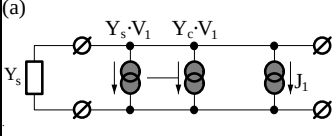
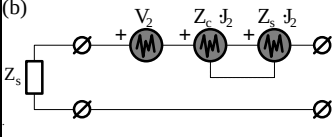
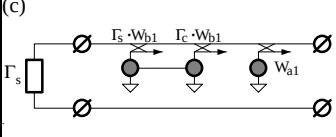
<i>Excess noise levels</i>	
(a)	 $\begin{aligned} S_i &= Y_s + Y_c ^2 \cdot S_{v1} + S_{J1} \\ S_i &= S_j + Y_s ^2 \cdot S_v + Y_s' \cdot S_{v,j} + Y_s' \cdot S_{Jv} \\ S_i &= S_j + Y_s ^2 \cdot S_v + 2 \cdot \text{re}\{Y_s' \cdot S_{Jv}\} \\ S_i &= S_j + Y_s ^2 \cdot S_v + 2 \cdot \text{re}\{Y_s\} \cdot \text{re}\{S_{Jv}\} + 2 \cdot \text{im}\{Y_s\} \cdot \text{im}\{S_{Jv}\} \\ S_i &= [Y_s, 1] \cdot C_A \cdot [Y_s, 1]' \end{aligned}$
(b)	 $\begin{aligned} S_u &= Z_s + Z_c ^2 \cdot S_{v2} + S_{V2} \\ S_u &= S_v + Z_s ^2 \cdot S_j + Z_s' \cdot S_{Jv} + Z_s' \cdot S_{V,j} \\ S_u &= S_v + Z_s ^2 \cdot S_j + 2 \cdot \text{re}\{Z_s' \cdot S_{V,j}\} \\ S_u &= S_v + Z_s ^2 \cdot S_j + 2 \cdot \text{re}\{Z_s\} \cdot \text{re}\{S_{V,j}\} + 2 \cdot \text{im}\{Z_s\} \cdot \text{im}\{S_{V,j}\} \\ S_u &= [1, Z_s] \cdot C_A \cdot [1, Z_s]' \end{aligned}$
(c)	 $\begin{aligned} S_w &= \Gamma_s + \Gamma_c ^2 \cdot S_{b1} + S_{a1} \\ S_w &= S_a + \Gamma_s ^2 \cdot S_b + \Gamma_s' \cdot S_{b,a} + \Gamma_s' \cdot S_{a,b} \\ S_w &= S_a + \Gamma_s ^2 \cdot S_b + 2 \cdot \text{re}\{\Gamma_s' \cdot S_{a,b}\} \\ S_w &= S_a + \Gamma_s ^2 \cdot S_b + 2 \cdot \text{re}\{\Gamma_s\} \cdot \text{re}\{S_{a,b}\} + 2 \cdot \text{im}\{\Gamma_s\} \cdot \text{im}\{S_{a,b}\} \\ S_w &= [1, \Gamma_s] \cdot C_T \cdot [1, \Gamma_s]' \end{aligned}$

Fig 7.8 Various related expressions for evaluating the device excess noise at specified source impedance. This excess noise equals the total equivalent input noise when the source impedances is cooled down to $T=0$ Kelvin.

Calculations of equivalent input noise using voltages and currents are known since 1955 by Rothe and Dahlke [712,713]. Additionally, they developed a method to represent the correlation with correlation impedances or -admittances as (positive and negative) circuit elements. This approach is not discussed here.

Wave methods have been discussed in the same period by Bauer and Rothe [714]. Penfield [718] proposed in 1962 to use noise waves, normalized to complex reference impedances, as will be discussed in the succeeding section 7.3.2. When Z_N is properly chosen ($Z_N = Z_{\text{opt}}$ see next section), the two noise waves become uncorrelated. This simplifies the noise wave analysis.

7.3.2. Conventional spot noise parameters for two-ports

The two-port noise-parameters discussed in section 7.3.1 originated from matrix concepts that are generally applicable to arbitrary multi-ports. This explains the regular nature of various expressions when using matrix noise parameters.

In the line of historical developments, other sets of noise parameters have been evolved. They have in common that they are associated with noise factor measurements as a function of source impedance. Those (spot) noise parameters are useful when designing

amplifier configurations with minimum noise factor, however, these design goals are not useful¹⁸ for wideband amplifiers.

To our opinion, conventional spot noise parameters are inconvenient for designing low-noise wideband feedback amplifiers. Nevertheless, it has been common practice to specify transistor noise parameters in this way. Furthermore, many publications relate their results to spot noise parameters. Therefore they are discussed in this section. Furthermore, conversion rules are provided between spot noise parameters and matrix noise parameters.

The basic concepts of the *noise factor* or *noise figure* were introduced by Burgess, Friss and North [705,706,707] in the 1940's and later adopted by IRE standards [817,818,821]. It is a quantity that relates the amplifier noise to the thermal noise of the associated source impedance. The factor F by which the total equivalent input noise is greater than the thermal noise of the source impedance is defined as the noise factor.

When S_i and S_u represent the spectral intensity of equivalent excess noise currents and voltages, and Y_s and Z_s are the source admittance and -impedance, then F is defined as:

$$F \stackrel{\text{def}}{=} \begin{cases} = \frac{S_i + 4kT \cdot \text{re}\{Y_s\}}{4kT \cdot \text{re}\{Y_s\}} \\ = \frac{S_u + 4kT \cdot \text{re}\{Z_s\}}{4kT \cdot \text{re}\{Z_s\}} \end{cases} \quad \text{definitions are restricted for } \text{re}\{Z_s\} > 0$$

$F-1 = \text{excess noise figure}$

The noise factor F is a quantity that changes with the source impedance, and therefore its value is meaningless when this source impedance is not specified. Goldberg [708] has shown in 1948 a possibility of noise factor reduction by mismatching impedances (noise tuning), an observation that was theoretically supported by Rothe and Dahlke [713]. In 1960 IRE standards recommended [818,819] two-port noise measurements, for extracting the minimum noise factor $F_{\min}$ and the associated optimal source impedance (Z_{opt} , Y_{opt} or Γ_{opt}). An additional parameter was required to facilitate the prediction of F for all source impedances (Z_s , Y_s or Γ_s) of interest. Initially an arbitrary additional parameter was chosen. A commonly used choice is (R_u or G_i) representing being an equivalent thermal noise resistance or conductance of noise voltage V or -current J as shown in figure 7.7a. In 1967, Lange [822] proposed a more fundamental parameter N . This is still an arbitrary parameter, however, it has the property that its value is invariant under lossless transformation (as holds for $F_{\min}$ too). This parameter is especially advantageous when the terminals of an intrinsic device are not directly accessible, e.g. due to the packaging of a chip.

The relations between matrix noise parameters and all these conventional spot noise parameters are shown in figure 7.9 and 7.10.

¹⁸ The noise figure is minimal for a specific 'optimal' source impedance. This requires an impedance transformation of the source impedance with a loss-free 'matching' network. This condition can easily be realized within a small frequency band using resonating circuitry such as a stub. On the other hand, minimum noise figure cannot be achieved for all frequencies within a wide frequency band.

C_A	C_T
$N \stackrel{\text{def}}{=} \frac{\sqrt{c_{A11} \cdot c_{A22} - \text{im}\{c_{A12}\}^2}}{4 \cdot k \cdot T}$ $F_{\min} = 1 + 2 \cdot N + \frac{\text{re}\{c_{A12}\}}{2 \cdot kT}$ $Y_{\text{opt}} = \frac{4kT \cdot N + j \cdot \text{im}\{c_{A12}\}}{c_{A11}}$	$N = \frac{\sqrt{(c_{T11} + c_{T22})^2 - 4 \cdot c_{T12} ^2}}{4 \cdot k \cdot T}$ $F_{\min} = 1 + 2 \cdot N + \frac{ c_{T11} - c_{T22} }{2 \cdot kT}$ $\Gamma_{\text{opt}} = \frac{2 \cdot c_{T12}}{\sigma \cdot 4kT \cdot N + c_{T11} + c_{T22}}$
$R_u \stackrel{\text{def}}{=} \frac{c_{A11}}{4kT} = \frac{N}{\text{re}\{Y_{\text{opt}}\}} = \frac{N}{\text{re}\{Z_0\}} \cdot \frac{ Z_0 + Z'_0 \cdot \Gamma_{\text{opt}} ^2}{1 - \Gamma_{\text{opt}} ^2}$ $G_i \stackrel{\text{def}}{=} \frac{c_{A22}}{4kT} = \frac{N}{\text{re}\{Z_{\text{opt}}\}} = \frac{N}{\text{re}\{Z_0\}} \cdot \frac{ 1 - \Gamma_{\text{opt}} ^2}{1 - \Gamma_{\text{opt}} ^2}$	
$Z_{\text{opt}} = 1/Y_{\text{opt}} = \frac{Z_0 + Z'_0 \cdot \Gamma_{\text{opt}}}{1 - \Gamma_{\text{opt}}}$ $Y_{\text{opt}} = 1/Z_{\text{opt}} = \frac{1 - \Gamma_{\text{opt}}}{Z_0 + Z'_0 \cdot \Gamma_{\text{opt}}}$ $\Gamma_{\text{opt}} = \frac{Z_{\text{opt}} - Z_0}{Z_{\text{opt}} + Z'_0} = \frac{1 - Y_{\text{opt}} \cdot Z_0}{1 + Y_{\text{opt}} \cdot Z'_0}$	

Fig 7.9 Definitions of various spot noise parameters, normalized to arbitrary reference impedance (including negative complex values, differently for each port).

$C_A = \begin{bmatrix} c_{A11} & c_{A12} \\ c_{A21} & c_{A22} \end{bmatrix} = kT \cdot (F_{\min} - 1) \cdot \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix} + \frac{4kT \cdot N}{\text{re}\{Y_{\text{opt}}\}} \cdot \begin{bmatrix} 1 \\ -Y_{\text{opt}} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -Y_{\text{opt}} \end{bmatrix}'$ $C_A = \begin{bmatrix} c_{A11} & c_{A12} \\ c_{A21} & c_{A22} \end{bmatrix} = kT \cdot (F_{\min} - 1) \cdot \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix} + \frac{4kT \cdot N}{\text{re}\{Z_{\text{opt}}\}} \cdot \begin{bmatrix} -Z_{\text{opt}} \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -Z_{\text{opt}} \\ 1 \end{bmatrix}'$ $C_T = \begin{bmatrix} c_{T11} & c_{T12} \\ c_{T21} & c_{T22} \end{bmatrix} = \sigma \cdot kT \cdot (F_{\min} - 1) \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + \frac{\sigma \cdot 4kT \cdot N}{1 - \Gamma_{\text{opt}} ^2} \cdot \begin{bmatrix} +\Gamma_{\text{opt}} \\ 1 \end{bmatrix} \cdot \begin{bmatrix} +\Gamma_{\text{opt}} \\ 1 \end{bmatrix}'$
$F(Y_s) = F_{\min} + \frac{N}{\text{re}\{Y_{\text{opt}}\}} \cdot \frac{ Y_s - Y_{\text{opt}} ^2}{\text{re}\{Y_s\}} = F_{\min} + R_u \cdot \frac{ Y_s - Y_{\text{opt}} ^2}{\text{re}\{Y_s\}}$ $F(Z_s) = F_{\min} + \frac{N}{\text{re}\{Z_{\text{opt}}\}} \cdot \frac{ Z_s - Z_{\text{opt}} ^2}{\text{re}\{Z_s\}} = F_{\min} + G_i \cdot \frac{ Z_s - Z_{\text{opt}} ^2}{\text{re}\{Z_s\}}$ $F(\Gamma_s) = F_{\min} + \frac{4 \cdot N}{1 - \Gamma_{\text{opt}} ^2} \cdot \frac{ \Gamma_s - \Gamma_{\text{opt}} ^2}{1 - \Gamma_s ^2} = F_{\min} + \frac{4 \cdot R_u / Z_0}{ 1 + \Gamma_{\text{opt}} ^2} \cdot \frac{ \Gamma_s - \Gamma_{\text{opt}} ^2}{1 - \Gamma_s ^2}$
$F(Y_s) = 1 + \frac{1}{4kT \cdot \text{re}\{Y_s\}} \cdot [+Y_s, 1] \cdot C_A \cdot [+Y_s, 1]'$ $F(Z_s) = 1 + \frac{1}{4kT \cdot \text{re}\{Z_s\}} \cdot [1, +Z_s] \cdot C_A \cdot [1, +Z_s]'$ $F(\Gamma_s) = 1 + \frac{1}{kT \cdot (1 - \Gamma_s ^2) } \cdot [1, -\Gamma_s] \cdot C_T \cdot [1, -\Gamma_s]'$

Fig 7.10 Relations between spot noise parameters and matrix noise parameters and noise figure. The validation of these relations are restricted to $\text{re}\{Y_s\} > 0$,

Penfield [718] proposed in 1962 to use waves normalized to Z_{opt} as reference impedance. Such a choice simplifies various relations since Γ_{opt} becomes zero, matrix parameter c_{T12} becomes zero, and the noise waves in figure 7.7b become uncorrelated. This can easily be verified from the relations in figure 7.9.

A general accepted way of specifying spot noise parameters is lacking, although an increasing number of manufacturers specify their semiconductor devices in terms of: $[F_{\text{min}}, \text{re}\{\Gamma_{\text{opt}}\}, \text{im}\{\Gamma_{\text{opt}}\}, R_{\text{f}}/Z_0]$. Other manufacturers specify in terms of noise circles, which requires graphical extraction of the spot noise parameters from these plots. Some authors use the factor $(4 \cdot N)$ as noise parameter instead of (N) , while referring to the original definition of Lange [822]. Other authors use the *noise measure* M , defined by Haus and Adler [715,716], as one of the spot noise parameters. This illustrates how confusing it is to intercompare the noise of various devices.

7.3.3. Autonomous noise parameters for two-ports.

So far, two groups of two-port noise parameters have been discussed. The matrix- as well as the spot noise parameters are fully capable of representing measured results on arbitrary noisy two-ports. This approach is effective for computer assisted circuit analysis, however, it is inconvenient for circuit synthesis¹⁹.

The aim of section 7.3.3 is to find a parameter representation of noisy two-ports, using (1) uncorrelated virtual noise sources characterized by (2) simple spectra over the full frequency band of interest. This might simplify the synthesis of noisy circuits and the development of transistor noise models, similarly to the concept of *two-port virtual circuit parameters* as introduced in section 2.3.2. We refer to the intensity spectra of these virtual noise sources as *autonomous noise parameters*²⁰.

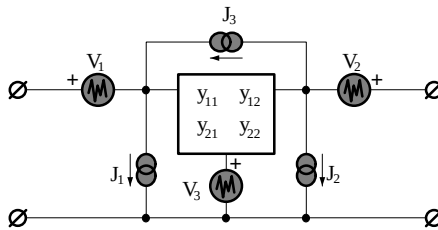


Fig 7.11 Autonomous noise parameters represent the intensity spectra of the uncorrelated noise sources in this circuit.

Autonomous noise parameters are related with a chosen topology, for instance the topology in figure 7.11. This topology is characterized by the fact that all virtual noise sources are extended to the output nodes of an inner noiseless blackbox. When the noise levels of two virtual noise sources are given (or chosen) the levels of the remaining four sources are unique.

¹⁹ The purpose of analysis is to find the properties of a circuit when it has been designed. Circuit simulations as well as measurements are adequate methods. The purpose of synthesis is to find a circuit that meets a set of pre-defined properties. This is the reverse problem.

²⁰ The term *autonomous* is chosen to emphasize that the virtual noise sources are to be *uncorrelated*.

The topology of figure 7.11 is not the only one to be considered, although other topologies have not been investigated. Alternative topologies can be based on a virtual circuit representation of a noiseless blackbox, that include internal (virtual) nodes. This facilitates the extension of virtual noise sources, connected with external as well as internal virtual nodes. Finding the most convenient virtual circuit topology for a device of interest is a matter of trial and error in practice.

Extraction algorithm.

All six noise sources of the topology in figure 7.11 are uncorrelated by definition, and the two-port transfer parameters are assumed to be known. Six noise levels are to be extracted from four (matrix) noise parameters, and therefore a unique solution does not exist. When the values for two of the six are chosen, for instance $S_{v3}=0$ and $S_{j3}=0$, then the remaining values can be extracted uniquely from measured matrix noise parameters. Let C_Y be the current correlation matrix of a two-port, and let Y be the associated two-port admittance parameters. To evaluate C_Y the noise currents can be inserted as discussed for circuit simulators in section 7.2.4. The noise voltages can be inserted using similar techniques and using the voltage to current transformation rule as summarized in figure 7.6. The extracted parameters $[S_{j1}, S_{j2}, S_{v1}, S_{v2}]$ from the measured C_Y and the chosen values $[S_{j3}, S_{v3}]$ are:

$$C_Y = S_{j1} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + S_{j2} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + S_{j3} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + S_{v1} \cdot Y \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot Y' + S_{v2} \cdot Y \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \cdot Y' + S_{v3} \cdot Y \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \cdot Y'$$

$$\begin{bmatrix} c_{Y11} & -1 & -(y_{11}+y_{12}) \cdot (y'_{11}+y'_{12}) \\ c_{Y21} & 1 & -(y_{21}+y_{22}) \cdot (y'_{11}+y'_{12}) \\ c_{Y12} & 1 & -(y_{11}+y_{12}) \cdot (y'_{21}+y'_{22}) \\ c_{Y22} & -1 & -(y_{21}+y_{22}) \cdot (y'_{21}+y'_{22}) \end{bmatrix} \cdot \begin{bmatrix} 1 \\ S_{j3} \\ S_{v3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & y_{11} \cdot y'_{11} & y_{12} \cdot y'_{12} \\ 0 & 0 & y_{21} \cdot y'_{11} & y_{22} \cdot y'_{12} \\ 0 & 0 & y_{11} \cdot y'_{21} & y_{12} \cdot y'_{22} \\ 0 & 1 & y_{21} \cdot y'_{21} & y_{22} \cdot y'_{22} \end{bmatrix} \begin{bmatrix} S_{j1} \\ S_{j2} \\ S_{v1} \\ S_{v2} \end{bmatrix}$$

$$\begin{bmatrix} S_{j1} \\ S_{j2} \\ S_{v1} \\ S_{v2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & y_{11} \cdot y'_{11} & y_{12} \cdot y'_{12} \\ 0 & 0 & y_{21} \cdot y'_{11} & y_{22} \cdot y'_{12} \\ 0 & 0 & y_{11} \cdot y'_{21} & y_{12} \cdot y'_{22} \\ 0 & 1 & y_{21} \cdot y'_{21} & y_{22} \cdot y'_{22} \end{bmatrix} \setminus \begin{bmatrix} c_{Y11} & -1 & -(y_{11}+y_{12}) \cdot (y'_{11}+y'_{12}) \\ c_{Y21} & 1 & -(y_{21}+y_{22}) \cdot (y'_{11}+y'_{12}) \\ c_{Y12} & 1 & -(y_{11}+y_{12}) \cdot (y'_{21}+y'_{22}) \\ c_{Y22} & -1 & -(y_{21}+y_{22}) \cdot (y'_{21}+y'_{22}) \end{bmatrix} \cdot \begin{bmatrix} 1 \\ S_{j3} \\ S_{v3} \end{bmatrix}$$

This left-hand matrix division always results in a real and unique solution. On the other hand, it will not necessarily yield positive values. An improper choice of $[S_{j3}, S_{v3}]$ may therefore result in solutions that have no meaning at all. The development of simple criteria for choosing $[S_{j3}, S_{v3}]$ is of interest for further investigation.

Extraction in practice

It is expected that the topology in figure 7.11 is convenient for the representation of noise in transistors. This is because that topology is closely related to simple physical noise models based on uncorrelated white noise sources. These physical models are shown in figure 7.12, and have often proven adequate for the design of [406] wideband amplifiers. When required, voltage noise sources can be added in series with the device terminals to represent thermal noise effects caused by series resistance.

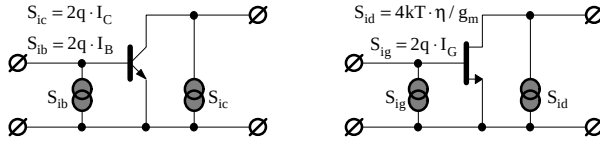


Fig 7.12 Simple noise models for transistors. Most noise sources originate from shot noise associated with bias currents or leakage currents. The channel noise in the FET is represented acting as if the transconductance g_m is providing thermal noise. The correction factor η is a value close to one, to be extracted from measurements. The FET model ignores the influence of induces gate noise because its contribution is of minor importance when designing wideband amplifiers [406]

Full two-port noise parameters of transistors are required to demonstrate the correctness of our proposition. In contrast to full two-port *transfer* parameters, these noise parameters are sparingly available and usually restricted to a few frequencies above 1 GHz. The lack of adequate measuring instruments has resulted in the development of an innovative setup for measuring two-port noise parameters, as described in section 8.5. Preliminary results [834], ranging from 10 to 500 MHz, indicate that the use of autonomous noise parameter is a promising technique. It is expected that this new approach simplifies the extraction of transistor noise models, similarly to the extraction of transistor transfer models from virtual circuit parameters

7.3.4. Conclusions

Two-port noise measurements can be specified in several different ways. Three classes of two-port noise parameters have been discussed in detail, as listed below:

- *Matrix noise parameters* are deduced from general applicable concepts. Various relations are therefore expressed in a regular way. This makes them most convenient for analyzing noise in wideband amplifiers.
- *Spot noise parameters* originate from historical developments on minimizing noise in very small frequency bands. Using these parameters in other applications result in unnecessary complicated expressions. Therefore they are inconvenient when designing wideband amplifiers.
- *Autonomous noise parameters* are promising new concepts, originated in this study. They may simplify the extraction of transistor noise models from two-port noise measurements. These parameters represent the spectral intensity of uncorrelated noise sources.

A general accepted way of specifying noise is lacking, although an increasing number of manufacturers is specifying their semiconductor devices in a corresponding spot noise parameter format. Most textbooks and publications are focused on some specific representation method. In some of them slightly different definitions on noise parameters are used.

This study resulted in generalized transformation rules between spot noise parameters and matrix noise parameters. They are extended to wave applications normalized to arbitrary reference impedance, including complex values with negative real part.

Chapter 8

Lightwave based electrical noise measurements

Noise measurements form an important aspect when designing circuits and systems. Spectral measurements of equivalent input noise are valuable for specifying linear circuit performance, and the circuit designer may use them to improve the individual circuits before they are combined to a system. When required, the system designer may use these spectral measurements for predicting bit-error rates when these circuits are used in future systems.

Spectral noise measurements on electrical receivers, related to 50 Ω sources, are relatively simple when using calibrated electrical noise sources. This does not hold for *lightwave* receivers since lightwave noise sources are currently (1994) not commercially available.

This work resulted in novel methods for measuring the spectral noise of electrical amplifiers and lightwave receivers. In addition, these lightwave methods have significant advantages when applied to electrical noise measurements, compared to methods using conventional 50 Ω noise sources.

Spectral measurements, related to 50 Ω sources, are valuable for specifying circuit performance. Their value will vary with the circuit topology and the choice of device type and device biasing. As a result, spectral noise measurements, related to 50 Ω sources, are inadequate for specifying transistors and other circuit devices.

Spectral *noise parameters* of the individual devices are more suitable to predict the circuit noise for an arbitrary circuit topology. They are essential for reliable noise analysis in a simulator, and useful to improve the circuit design. Furthermore, measured noise parameters are crucial to develop new noise models of devices.

Synthesis of wideband noise-optimal amplifiers requires reliable noise models of transistors and other devices. Simple transistor noise models are commonly used [406] to successfully minimize noise at relatively low frequencies. When designing wideband lightwave receivers, we observed a disagreement between noise measurements and circuit noise analysis using these simple noise models. It requires a set of transistor noise parameters at various frequencies to develop improved noise models.

Manual methods for extracting device noise parameters using noise sources and impedance tuners are well-known and recommended since 1960 by IRE standards [818]. These measurements are quite laborious, and only recently some fully automated instruments have come to the market, mainly for microwave frequencies. As a result, complete sets of transistor noise parameters are sparingly specified, and if they are they are commonly restricted to a few distinct frequencies above 1 GHz.

This work resulted in novel methods for measuring two-port noise parameters, to facilitate future developments on improved transistor noise models. These methods are

an advantage when measuring below 1 GHz, compared to conventional methods using 50 Ω noise sources cascaded by two-port impedance tuners.

This chapter describes various noise measurement methods on lightwave receivers, electrical amplifiers and transistors, using novel lightwave-based principles. It introduces new devices, calibration methods and measurement methods.

The measurement techniques proposed here rely on a novel lightwave device, a *synthetic noise generator*, providing white noise when illuminating PIN photo-diodes. This chapter is restricted to the application of this white noise source; the lightwave source is described in detail in chapter 9.

The highlights of this chapter, which are developed during the work on this thesis, are:

- Development of a new device, a *lightwave noise-tee* [802], applicable for calibration of synthetic noise, as well as generation of electrical noise. It is a multi-functional box, that has been used for various types of noise measurements.
- Development of a novel *multi-level noise source*, using a synthetic noise generator in combination with a lightwave noise-tee [802]. This combination has many advantages, in comparison to commonly used 50 Ω noise sources.
- Development of various *calibration* methods for synthetic noise, including attractive methods based on shot-noise calibration. The proposed shot-noise calibration is associated with high inherent accuracy (<1dB) obtained from simple dc-measurements. This is a significant advantage when initial calibration with primary noise standards is omitted.
- Development of new methods to *transform the calibration* of noise to arbitrary reference planes, using mathematical halving of noise-tees. The transformation method facilitates noise measurements using arbitrary source impedances.
- Development of a novel method to measure equivalent input noise of *lightwave receivers* [801, 803: patent pending]. This method facilitates noise measurements valid at the *optical* input of the receiver (usually the receiver noise is specified at an internal electrical reference plane behind the photo detector).
- Development of a novel algorithm to extract *equivalent input noise* of an optical or electrical device under test. It facilitates *multi-level* noise measurements to improve the measurement accuracy of noise measurements at comparative measurement time. This means that more reference levels are used than the commonly used 'hot' and 'cold' noise levels. The multi-level noise source is matched to the transmission line that links the source with the device under test.
- Development of a novel algorithm that *generalizes* the two-port *noise parameter* extraction of an electrical device under test, with known input impedance. The algorithm holds for any combination of noise sources, each with arbitrary effective 'temperature' and impedance. When more than five different sources are used, the algorithm extracts the most plausible solution, by minimizing the *relative* error. The algorithm includes the application of *multi-level* noise measurements.

This chapter is restricted to spectral density measurements on noise. Other measurements, such as BER measurements, are not considered. Furthermore, correction for out of band effects, such as spurious response errors, are not discussed.

8.1. The art of measuring noise

Noise measurements can be performed as *direct* noise measurement as well as *ratio* noise measurement. Both methods results in *absolute* noise figures, although ratio-noise measurements are superior in accuracy and simplicity. Typical aspects of both methods are listed below:

- For *direct* measurements the output noise spectrum of an amplifier under test is measured with a calibrated selective voltmeter, such as a spectrum analyzer or a noise figure meter. Another type of measurement determines the gain of the amplifier. The input noise is reconstructed from the measured output noise by division by the measured gain. The source and load impedances of the measurement are kept equal to the source and load impedance of the application.
- For *ratio* measurements, a calibrated white noise source is connected to the input of an amplifier under test. The *relative* change in output noise spectrum is measured when the level of the calibrated noise source is varied (usually switched between 'high' and 'low'). The input noise is reconstructed from this relative change and from the known noise levels of the noise source. The *absolute* value of the amplifier output spectrum is irrelevant.

This section discusses both measurement methods, to illustrate the superiority of ratio noise measurements. They are the most convenient, because the measurement accuracy is mainly determined by the calibration accuracy of a (simple) noise source. The remainder of this chapter is mainly restricted to ratio measurements.

8.1.1. Pitfalls while measuring noise

High accuracy is hard to obtain for direct methods. Spectrum analyzers are convenient instruments for this purpose, however ± 2 dB amplitude uncertainties are common values. This accuracy holds for detection for harmonic signals and not for noisy signals. The accuracy is reduced for noise by built-in envelope detectors, logarithmic amplifiers and fast sweeping filters with variable resolution bandwidth. This can be explained as follows [804]:

- The built-in detector is usually not a true-rms detector, but is based on a much simpler top detector. When its detection level is calibrated for Gaussian distributed noise, it provides erroneous results when the statistic distribution of the input signal is non-Gaussian (we observed 0.2 dB difference when using our synthetic noise source, as described in section 9.2.4).
- The built-in logarithmic amplifier modifies the statistic distribution which prevents correct noise detection, even when an rms-detector is used [804]. This effect can be eliminated when the spectrum analyzer is used in true linear mode.
- Variation of the resolution bandwidth varies the detected noise level, while this level will not vary when detecting harmonic signals.

As a result, when a spectrum analyzer is used in log mode, and noise corrections are restricted to resolution bandwidth correction, the average noise is displayed approximately 2~2.5 dB **too low!** [804]

Ratio measurements are insensitive to scaling errors in the frequency selective voltage detection. Gain and resolution bandwidth measurements are not required, and mis-match errors will hardly decrease the overall accuracy. Therefore, the accuracy of ratio

measurements is mainly restricted to the accuracy of the calibrated noise source. Since this device is simple, it is easy to calibrate it with high accuracy; ± 0.1 dB are common values. As a result, ratio noise measurements are superior in accuracy and simplicity, compared to direct noise measurements

8.1.2. Accuracy limits when measuring noise

An increasing number of spectrum analyzers is equipped with (hidden) calibrated noise marker facilities, to simplify their use in direct noise measurements. These facilities perform the required error corrections automatically, and create the impression of accuracy. This is because the measurement accuracy for noisy signals with unknown statistic distribution is always lower than the accuracy for harmonic signals. These limitations are more significant when logarithmic receivers are used than when linear receivers are used. This is discussed below and in section 9.2.4.

- *Accuracy limits in logarithmic receivers (as in many spectrum analyzers)*

The spectrum analyzer is primarily intended for detecting signal components of which the spectral width is small compared to the resolution bandwidth. An additional noise marker corrects for systematic errors [804], and its calibration is commonly based on natural (Gaussian distributed) noise.

To simplify matters a noise marker calibration is assumed with equal accuracy performance as applies to harmonic signals. The noise marker calibration corrects for errors originating from the use of an envelope detector, to simulate an rms detector. When the statistic distribution of noisy signals differs from Gaussian distributions then this correction will fail. This yields an additional uncertainty, which makes the overall accuracy for noise detection lower than for harmonic signal detection.

Furthermore, replacement of the envelope detector by a true-rms detector will not always resolve this problem with the noise marker accuracy. A logarithmic amplifier compresses the peaks of noisy signals and modifies their statistic distribution before it is detected. As a result, the combination of true-rms detection with logarithmic amplification remains sensitive to variations in statistic distribution.

- *Accuracy limits in linear receivers (as in many noise figure meters)*

To improve the measurement accuracy for noisy signals, commonly used noise figure meters (e.g. HP8970a) are implemented as linear receiver¹, instead of logarithmic receivers. This makes them more dedicated to noise measurements but slower responding on large signal variations.

Although these instruments are dedicated to noise measurements, it is not a matter of course that they are equipped with *true-rms* detection (HP8970a uses an envelope detector). This increases the overall measurement uncertainty, compared to true-rms detection, when non-Gaussian distributed signals are detected.

A remaining noise accuracy problem arises from the spectral width of noisy signals. Frequency selective receivers are sensitive to out-of-band signals (spurious

¹ These instruments are equipped with switched attenuators to scale the input signal within the linear detection range of the noise detector. Since this requires an iterative adjustment when the signal level changes, these linear receivers are slower responding than receivers with logarithmic converters.

response). This results from inter-modulation by non-linear distortion and from insufficient image rejection. The wider the spectral width is, the more signal is erroneously detected. This inaccuracy increases with the magnitude of the noisy signal. As a result, spectral noise detection suffers from additional accuracy reduction, compared to spectral detection of harmonic signals.

The use of additional filtering (preselection), before heterodyne mixing, will improve the accuracy, however this is commonly not a standard feature in spectrum analyzers.

In general, ratio noise measurements are less sensitive to the above mentioned accuracy limits. This holds especially when the statistical distribution, spectral width and spectral level of the calibrated noise source equals the noise of the device under test.

8.1.3. Basic definitions

In ratio noise measurements, the output noise power of the device under test varies with the noise level of the calibrated noise source. It is often convenient to relate these noise levels to thermal noise. The terms cold noise, excess-noise and hot noise are defined as follows:

- *Cold noise* is the output noise of the device under test, when the noise source is switched 'off'. The lowest noise level of this source equals the thermal noise of its source impedance at room temperature, for instance $R=50\ \Omega$ and $T=290\text{ K}$. The intensity spectrum of this noise level will be referred as $S_c(\omega)$.
- *Excess noise* is what an activated noise source adds to the cold noise level. This level will be referred as $S_e(\omega)$.
- *Hot noise* is the output noise of the device under test when the noise source is switched 'on'. It is the combination of cold-noise and excess-noise. Since they are uncorrelated, the intensity spectrum of this noise is $S_h(\omega) = S_c(\omega) + S_e(\omega)$.

An important figure of merit of calibrated noise sources is the ratio of their excess noise level and cold noise level. The *excess noise ratio* (in dB) of a noise source is defined as: $ENR_{dB} = 10 \cdot \log(S_e/S_c) = 10 \cdot \log(S_h/S_c - 1)$.

Level S_h is the (hot) spectral intensity when the source is switched on, and S_c the (cold) spectral intensity when the source is switched off.

When the ENR of a source are specified and the cold noise level is a known reference level, the associated excess noise and hot noise levels can be extracted from these values. The cold noise level is usually the thermal noise of the source impedance, at the IRE *standard reference temperature*. according to IRE standards [817], this temperature is defined as $T=290\text{K}$. This is rather cold for convenient room temperature, however, it has been chosen because $(kT/q)=0.0250\text{V}$ (q =elementary charge).

Room temperature is usually 2% higher, and for that reason automated noise figure meters such as an HP8970a use $T=296.5\text{K}$ as default reference temperature.

8.2. Multi-level noise source with lightwave noise-tee

In order to carry out ratio noise measurements on electrical amplifiers, various white noise sources are commercially available. Most of these sources are provided with fixed output impedance (50Ω) and with fixed on and off noise levels.

The use of more than two calibrated noise levels is attractive to reduce random measurement errors by redundancy. The use of output impedances other than 50Ω is attractive when measuring noise levels below the thermal noise level of 50Ω resistors.

This study has resulted in the introduction [802] of a new lightwave device, called a *noise-tee*², for generating white electrical noise. It uses a lightwave synthetic noise generator [901,902,907], also resulted from this study. The noise-tee has attractive additional features, compared to commonly used 50Ω noise sources, including:

- Noise level in 'off'-state is significantly lower than thermal noise of 50Ω resistors.
- Noise level is variable over a wide dynamic range (ENR: 0-40dB, ore more).
- Relative accuracy of noise level is based on simple transfer measurements.
- Absolute accuracy of noise level is based on simple calibration at low frequencies
- Scaling accuracy of noise level is based on simple dc current measurements.
- Output impedance is variable over a wide dynamic range (up to $100\text{ k}\Omega/0.2\text{pF}$).
- Output impedance is insensitive to variation of noise level.
- Applicable in one-port and two-port configurations.

This subsection describes the proposed noise-tee and its performance.

8.2.1. Circuit diagram of a noise-tee

Figure 8.1 shows the basic circuit diagram of the noise-tee [802]. The noise-tee is a transmission line (e.g. microstrip or stripline), that is shunted in the middle by an illuminated PIN photo-diode. White noise is generated in this diode using a lightwave *synthetic noise generator* [901,902,907], a new lightwave device that is described in chapter 9. The PIN photo-diode is (externally) biased, and the dc photo current through the PIN photo-diode is (externally) sensed to detect the illumination power.

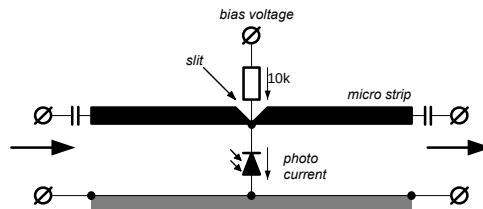


Fig 8.1 Basic circuit diagram of a lightwave noise-tee. A noise current is generated in a photo-diode that is shunted to a (microstrip) transmission line. The microstrip slit compensates for the diode capacitance.

² The name (lightwave) *noise-tee* is proposed in [802], and inspired to the name 'bias-tee'. A noise-tee injects a noise current in a transmission line while a bias-tee injects a dc-bias current.

As a matter of course, a simple one-port construction with a PIN photo-diode, shunted by a 50Ω resistor, is adequate to implement a 50Ω noise source. This situation is similar to our two-port configuration of figure 8.1, with an external 50Ω resistor on one of its ports. Nevertheless, the symmetrical two-port construction increases the flexibility of the device, as will be demonstrated in this chapter. It enables an external modification of source impedance.

The high impedance of the photo-diode causes a minimal degradation of the propagation performance of the transmission line. The most annoying effect is the parasitic capacitance of the PIN photo-diode. Nevertheless, capacitance values below 0.2pF are feasible with naked chips. A microstrip slit (see figure 8.1) is effective in compensating these parasitic capacitances.

All noise-tee experiments in this chapter are based on a 50Ω microstrip construction on 1.5 mm glass-epoxy board, 4 cm long, and based on a 1 pF photo-diode with 2.5 nH serial inductance. The overall responsivity of this construction is flat up to 1.5 GHz .

8.2.2. Variable noise source based on a matched noise-tee configuration

One of the noise-tee applications is its usage in a stand-alone *electrical* noise source. Figure 8.2 shows the associated (matched) configuration. The lightwave output signal of a synthetic noise generator is fed to the noise-tee using a variable optical attenuator and an optical fiber. The illumination generates a white noise current in the PIN photo-diode, proportional with the illuminated optical power. The photo current is externally measured with a pico-ampere meter to detect this illumination power.

The left side of the noise-tee is externally matched to the characteristic impedance Z_0 of the internal transmission line of the noise-tee (see figure 8.1). As a result, the output impedance at the right side becomes Z_0 too, e.g. $Z_0=50\Omega$. Furthermore, it prevents the output level of the source from being frequency dependent when the internal noise-tee transmission lines have non-zero lengths.

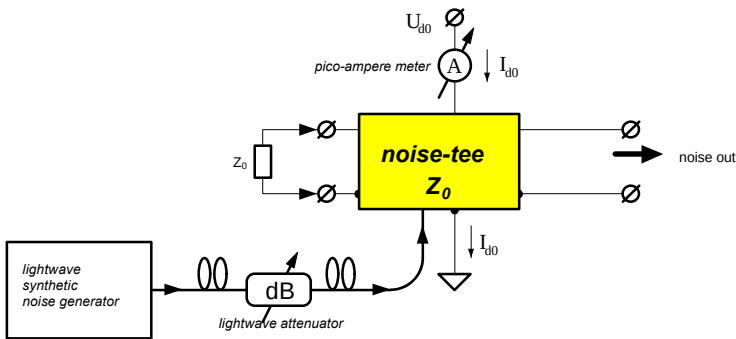


Fig 8.2 Matched noise-tee configuration to implement an electrical noise source with variable output level. The ratio between photo current I_{d0} and spectral current density ($\mathcal{S}_{i_e}$) of the excess-noise is insensitive to optical power variations.

Lightwave synthetic noise generators (see chapter 9) may generate noise that is white over several hundreds of gigahertz [902]. The magnitude of the noise-tee frequency response, from optical input port to electrical output port, is independent of the transmission line length when both sides are loaded with Z_0 . As a result, the lightwave frequency response of the noise-tee is mainly limited by the bandwidth of the PIN photo-diode. A bandwidth of more than 10 GHz is feasible.

The lightwave frequency response of the noise-tee can be measured simply and accurately with a heterodyne setup. Therefore, it is adequate to calibrate the noise current in a small frequency band, and use the lightwave frequency response to extrapolate this value to other frequencies. In many applications, this extrapolation is simple because PIN photo-diodes have a flat frequency response over a wide frequency interval.

8.2.3. Output noise level of a matched noise-tee configuration

One of the most attractive features of a matched noise-tee configuration (figure 8.2) is that the output noise level is continuously variable, when using an optical attenuator. Since the minimum output noise level is restricted by the thermal noise of the internal and external resistors, it is convenient to relate the various noise contributions to thermal noise. Usually, the terms cold noise, excess-noise and hot noise apply, as defined in section 8.1.

- *Cold* noise is the output noise of the configuration in figure 8.2, without illumination. The intensity spectrum of this current will be referred as $S_{ic}(\omega)$.
- *Excess-noise* is the synthetic noise current flowing through the photo-diode. The intensity spectrum of this current will be referred as $S_{ie}(\omega)$.
- *Hot* noise is the total output noise of an illuminated noise-tee: The intensity spectrum of this current is $S_{ih}(\omega) = S_{ic}(\omega) + S_{ie}(\omega)$.

The cold noise mainly originates in the external (50Ω) load (in figure 8.2 on the left side of the noise-tee). When this load is removed, the cold noise is significantly reduced, and is limited by the thermal noise originating from the internal bias resistors (e.g. $10k\Omega$, or higher). When this external load is cooled, the cold noise level reduces, without affecting the output impedance and excess-noise level. This might be an advantage for dedicated microwave measurements on (cooled) ultra low-noise amplifiers.

The excess-noise originates from the illumination by the lightwave synthetic noise generator. The amplitude spectrum of this current ($\sqrt{S_{ie}}$) is proportional to the illuminating power.

The responsivity of the PIN photo-diode is independent of the optical power, which means that the *ratio* between dc photo current (I_{d0}) and the spectral current density ($\sqrt{S_{ie}}$) of the excess-noise is a constant over a wide dynamic range. We verified this experimentally in reference [802].

This property is of great value when the excess-noise level is varied with an optical attenuator. Once calibrated at a specific reference noise level S_{ie0} , the excess-noise of the setup in figure 8.2 is accurately scaled to an arbitrary noise level $S_{ieX} = (I_{dX}/I_{d0})^2 \cdot S_{ie0}$ when the dc photo current I_d is measured with a simple (pico)-ampere meter.

The maximum excess-noise of a noise-tee configuration is significantly higher than for calibrated noise sources such as an HP346c (50Ω , $\text{ENR}=13\text{dB}$, 10MHz to 26GHz). This is because synthetic noise generation is a power efficient process (section 9.2.3). For example, assume a lightwave synthetic noise generator that provides synthetic noise over $B=50\text{GHz}$ bandwidth using a 1mW laser at 1300nm . Under these circumstances, the mean (dc) photo current in the noise-tee is roughly $I_{d0}=0.5\text{mA}$. Using the theory on synthetic noise generation, as discussed in section 9.2.3 and 9.3.2, the calculated excess-noise current density is approximately $\sqrt{S_{ie}}\approx 1.6\text{nA}/\sqrt{\text{Hz}}$. This value is more than 87 times higher than the thermal noise current in a 50Ω resistor $\sqrt{S_{ie}}\approx 18\text{pA}/\sqrt{\text{Hz}}$, which demonstrates the high excess-noise level that is available ($\text{ENR}\approx 39\text{dB}$).

8.2.4. Output impedance variation of a noise-tee configuration

Another attractive feature of a matched noise-tee configuration (figure 8.2) is that the output impedance is insensitive to variations of the output noise level, over a wide dynamic range. Figure 8.3 demonstrates this in competition with an HP346c noise source.

The change in output impedance is negligible for the proposed noise-tee, while an HP346c noise source is liable to $\pm 10\%$ impedance variations when it is switched between on and off. This impedance variation may result in measurement errors when the noise contribution and the gain of an amplifier under test vary with the source impedance.

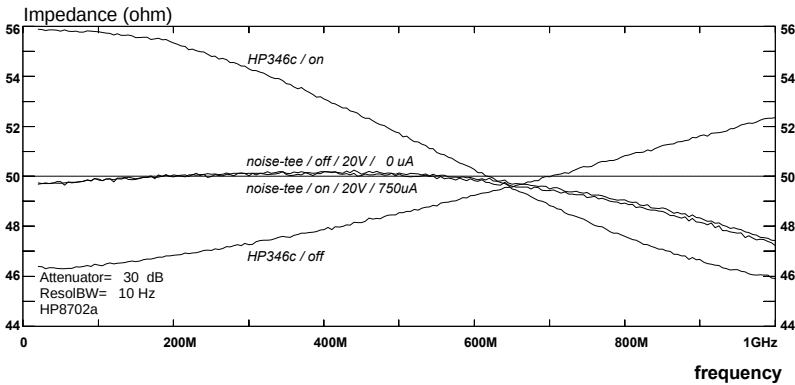


Fig 8.3 Output impedance measurements on a matched noise-tee configuration and a conventional noise source (HP346c), using 201 frequency points. One port of the noise-tee is loaded with 50Ω . The output impedance of the other port is insensitive to variations of noise level, while variations of more than $\pm 10\%$ are observed for conventional noise sources.

As a matter of course, impedance variations of a noise source can be reduced by an additional matched attenuator or isolator. The attenuator reduces the excess-noise ratio, which restricts the measurement accuracy when the ENR is relatively small. The isolator reduces the usable noise bandwidth too, and thus the applicability of the noise source.

Furthermore it is observed that the output impedance of the noise-tee is (nearly) insensitive to variations of the noise-tee bias voltage, when it ranges from 5V to 20V. Nevertheless, the highest precision (e.g. less than 0.01 dB variation) will be obtained when the PIN photo-diode voltage is sensed and adjusted to a constant value with a dc feedback loop.

Another attractive feature is that the output impedance is variable over a wide range. This is performed by replacing the external load with a one-port impedance tuner. Fifty ohm calibrated noise sources require cascading with two-port tuners for similar impedance variation. This configuration blocks or shorts the signal flow when relative high or low impedances are required, which illustrates the advantage of our noise-tee configuration.

Figure 8.4 demonstrates how this is used in a setup that measures the input noise of an amplifier under test, at specified source impedance. The impedance tuner is adjusted in such a way that the output impedance of the noise-tee has the required source value. Note that this output impedance is frequency dependent, due to the internal transmission line of the noise-tee. As a result, the realization of the desired output impedance is often restricted to small frequency bands.

The output impedance of the noise-tee is variable over a wide range and is limited by the photo-diode capacitance, its bias network, and the internal transmission line. The highest values are obtained when photo-diode and impedance tuner are integrated with the amplifier chip (e.g. 0.2pF//100k Ω). To obtain a high output impedance at relatively low frequencies, e.g. below 100 MHz, the dimension of the transmission line in the noise-tee is of minor importance.

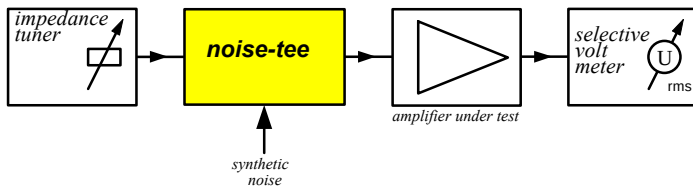


Fig 8.4 Mis-matched noise-tee configuration for noise measurements on an amplifier under test, at specified source impedance. The construction of this configuration is simpler than the combination of a two-port impedance tuner and a one-port noise source.

Another attractive feature of our noise-tee is that the mis-matched configuration of figure 8.4 facilitates the realization of 100% reflective sources (e.g. using an impedance tuner with infinite impedance). Examples are perfect shorts (0 Ω) perfect opens ($\infty \Omega$) and offset shorts and opens. When not in use, our noise-tee is equivalent with a pure transmission line, with no need to remove it.

This is a serious advantage, compared to conventional two-port impedance tuners. When a loss-free two-port tuner becomes 100% reflective, all available noise power from the noise source is blocked. This does not hold when source and tuner are reversed in position, however this requires an additional circulator.

Output impedances that are significantly different from 50Ω are attractive when measuring noise currents below the thermal noise level of 50Ω resistors. An example is the input noise current of an FET, especially at relatively low frequencies.

Commercially available setups for transistor noise parameter measurements, use 50Ω noise sources cascaded by a two-port impedance tuner. Since their output impedance is fixed, they require two-port impedance tuners to transform this impedance value to the desired value. At microwave frequencies, impedance transformation can be realized easily with resonating networks, however relatively low frequencies require other solutions, such as winded transformers.

As a result, the proposed setup in figure 8.4 is superior in simplicity, in competition with 50Ω calibrated noise sources and two-port impedance tuners.

8.2.5. Conclusions

In conclusion, it has been demonstrated that noise-tees, introduced in [802], have attractive additional features, compared to commonly 50Ω noise sources. Their noise level is easily varied over a wide dynamic range, while the associated noise level is accurately scaled from a calibrated reference value using simple dc-current measurements.

Furthermore the output impedance is mainly determined by an external impedance, and independent of the generated noise level. The mis-matched configuration of the noise-tee facilitates impedance tuning with one-port impedance tuners, while conventional methods require two-port tuners or circulators. Moreover, the mis-matched configuration facilitates a simple realization of 100% reflective sources. This is an advantage, compared to conventional two-port impedance tuners.

When not in use, our noise-tee is equivalent with a pure transmission line, with no need to remove it.

8.3. Calibration of synthetic noise

The measurement of noise, using white noise sources, requires a precise specification of the output noise level of the noise source. One way to perform this is to measure the spectral current density of the excess-noise current with an accurate frequency selective true-rms meter. More convenient methods use thermal noise or shot-noise as reference, because their values are based on fundamental physical constants. Thermal noise currents and shot-noise currents have the following well-known levels (single sided intensity spectra):

$S_i = 4kT/R$	$= \text{level of thermal noise current of resistor } R$
$S_i = 2q \cdot I_{dc}$	$= \text{level of shot-noise current of dc-current } I_{dc}$
$k = (1.380662 \pm 0.000044) \cdot 10^{-23} \text{ [J/K]} \text{ Boltzmann constant}$ $q = (1.6021892 \pm 0.0000046) \cdot 10^{-19} \text{ [C]} \text{ Elementary charge}$ $T = \text{absolute temperature; } 290\text{[K]} \text{ is the standard room temperature}$ $S_i = \text{(single sided) intensity spectrum of current [A}^2\text{/Hz]}$	

Thermal noise calibration may provide accurate results, however it requires a complex setup to heat and cool (50Ω) resistors. The use of calibrated noise sources, such as an HP346c, is preferred as intermediate step when their calibration is traceable to a noise standard. Shot-noise calibration is superior in simplicity when lightwave (synthetic) noise sources are available.

This section describes novel calibration methods for lightwave synthetic noise using (1) calibrated noise sources or (2) shot-noise generated in illuminated PIN photo-diodes. The proposed methods are applicable to a wide range of noisy signals of lightwave origin, including RIN (relative intensity noise) of lasers and LED's.

8.3.1. Definition of (spectral) noise-current ratio for synthetic noise

Synthetic noise originates from the illumination of a PIN photo-diode, using a lightwave synthetic noise generator. The intensity spectrum S_i of the generated (random) photo current $I(t)$ is frequency independent over a wide frequency band, and this property is referred as white noise. This noise is the uncorrelated combination of synthetic noise and shot-noise:

$$\begin{aligned}
 I_{dc} &= \text{dc-current; mean value of the random photo-current } I(t) \\
 S_{i,\text{shot}} &= \text{shot-noise level } (2q \cdot I_{dc}) \text{ associated with photo-current } I_{dc} \\
 S_{i,\text{synth}} &= \text{synthetic noise level, superposed on the photo-current } I_{dc} \\
 S_{i,\text{tot}} &= \text{total (excess) noise level: } S_{i,\text{synth}} + S_{i,\text{shot}} = S_{i,\text{synth}} + 2q \cdot I_{dc}
 \end{aligned}$$

DC photo current I_{dc} is proportional to illuminated power, and the same applies for the amplitude of all ac terms superposed on that current. As a result, the ratio between amplitude spectrum ($\sqrt{S_{i,\text{synth}}}$) and the mean value (I_{dc}) of the photo-current is constant when the illumination is varied (see lightwave attenuator in figure 8.2). We define:

$$\mu \stackrel{\text{def}}{=} \frac{\sqrt{S_{i,\text{synth}}}}{I_{\text{dc}}} = \frac{\sqrt{S_{i,\text{tot}} - 2q \cdot I_{\text{dc}}}}{I_{\text{dc}}} = (\text{spectral}) \text{ noise current ratio}$$

Calibration of synthetic noise is focused on the measurement of this constant μ . In this text, this constant will be referred as (spectral) *noise-current ratio* (NCR)³.

In [802], it has been demonstrated that this ratio is constant over a wide dynamic range. As a result, specification of μ and the measurement of I_{dc} facilitates the calculation of the noise level, using: $S_{i,\text{tot}} = (\mu \cdot I_{\text{dc}})^2 + 2q \cdot I_{\text{dc}} \approx (\mu \cdot I_{\text{dc}})^2$.

In most practical situations, the shot-noise $2q \cdot I_{\text{dc}}$ is negligible relative to the synthetic noise level. The bandwidth B is another characteristic constant of a synthetic noise generator, because approximately 50% of B is usable for white noise purposes (see section 9.2.3). In section 9.3.2 it will be demonstrated that B and μ are closely related for a well-designed synthetic noise generator:

$$\mu \approx \frac{1}{\sqrt{2 \cdot B}}$$

This approximation illustrates that shot-noise effects can be ignored when $I_{\text{dc}} \gg 2q/\mu^2 \approx 4q \cdot B$. For a 50 GHz synthetic noise generator this means that $I_{\text{dc}} \gg 0.03 \mu\text{A}$. Since this condition is easily fulfilled, this text will further ignore this shot-noise side effect.

8.3.2. Calibration of synthetic noise with calibrated noise sources

Synthetic noise can be calibrated simply when it is injected in a matched noise-tee configuration. The generated photo current is exactly twice the output excess current when both electrical sides are equally matched to the characteristic impedance Z_0 of the noise-tee. The proposed calibration uses this property in a mis-matched noise-tee configuration, in which the load of one port is replaced by a calibrated noise source with equal output impedance Z_0 . Figure 8.5 shows the associated calibration setup, in which the lightwave synthetic noise is applied to the noise-tee as is illustrated in figure 8.2.

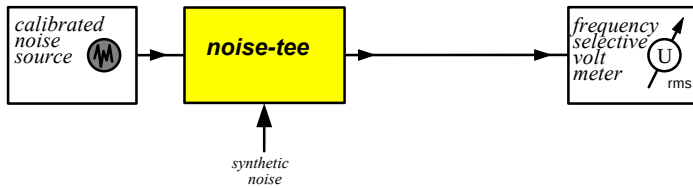


Fig 8.5 Noise-tee configuration for calibration of synthetic noise with a calibrated noise source.

³ Our definition is closely related to that of laser RIN (relative intensity noise). RIN is commonly specified in dB/Hz using $10 \cdot \log_{10}(\mu^2)$. We used another name for μ since intensity noise is quite different from synthetic noise. Typical values for laser intensity noise and synthetic noise are -155 and -110 dB/Hz respectively.

The calibration is performed in two steps:

- In the first calibration state the calibrated noise source is activated, while the noise-tee is not illuminated. Since the noise-tee is nearly loss-free, the frequency selective voltmeter will detect the 'hot' noise level, which is accurately specified. This will be referred as the *calibration (noise) level*.
- In the second calibration state the calibrated noise source is switched-off, while the illumination of the noise-tee is switched on. The illumination level is adjusted by the optical attenuator (see figure 8.2) to make the output noise level equal to the (hot) level of the first calibration state. As a result the synthetic noise current has become equal to the (specified) excess-noise current of the calibrated noise source. This will be referred as the *reference (noise) level*, and the associated dc photo current as the *reference (dc) current* $I_{d,ref}$.

Figure 8.6 shows the current flow for both calibration states.

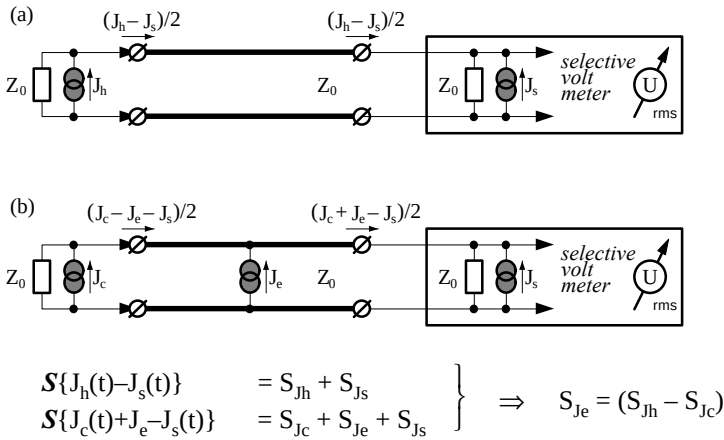


Fig 8.6 Current flow (time domain) of the various noise currents when (a) the source is on and the illumination is off, and (b) the reverse situation. In this model, the photo-diode capacitance is ignored. In our experimental noise-tee, this approximation is adequate up to 1 GHz.

The intensity spectrum S_{J_e} of the synthetic noise current $J_e(t)$ equals the excess-noise current of the calibrated noise source, when both calibration states yield equal noise levels. As a result, the reference level equals a specified calibration level. The (spectral) noise-current ratio μ is the ratio of the *reference level* $\sqrt{S_{i,ref}} = \sqrt{S_{J_e}}$ and the *reference current* $I_{d,ref}$.

Since the excess-noise current of the calibrated noise source is usually specified by an excess-noise ratio ξ (ENR), its value must be reconstructed using the standard room temperature ($T_c = 290K$ [817]) and the specified source impedance ($R_0 = 50\Omega$):

$$\mu \stackrel{\text{def}}{=} \frac{\sqrt{S_{i,ref}}}{I_{d,ref}} = \frac{\sqrt{S_{J_e}}}{I_{d,ref}} = \frac{\xi \cdot \sqrt{S_{J_c}}}{I_{d,ref}} = \xi \cdot \frac{\sqrt{4kT_c/R_0}}{I_{d,ref}} \quad \text{ENR} = 20 \cdot \log_{10}(\xi) \text{ [dB]}$$

For example, assume a calibrated noise source, with $R_0 = 50\Omega$ output impedance, $\xi = 13\text{dB}$ excess-noise ratio, and room temperature $T_c = 290[K]$. Furthermore assume that

the synthetic noise generates $I_{d0}=20\mu\text{A}$ dc photo current while it equals the specified excess-noise. Then the various quantities are:

$$\begin{aligned}\xi &= 13\text{dB} = 10^{13/20} \approx \sqrt{(19.95)} && \text{excess-noise ratio} \\ S_{Jc} &= 4kT_c/R_0 && \sqrt{S_{Jc}} \approx 18.1 \text{ pA}/\sqrt{\text{Hz}} && \text{cold noise} \\ S_{Je} &= \xi^2 \cdot S_{Jc} && \sqrt{S_{Je}} \approx 80.8 \text{ pA}/\sqrt{\text{Hz}} && \text{excess-noise} \\ S_{Jh} &= (\xi^2+1) \cdot S_{Jc} && \sqrt{S_{Jh}} \approx 82.8 \text{ pA}/\sqrt{\text{Hz}} && \text{hot noise} \\ \mu &= (\sqrt{S_{Je}})/I_{d0} \approx 4.0 \text{ (pA}/\sqrt{\text{Hz}})/(\mu\text{A}) \approx 1/\sqrt{(2 \cdot 30.6\text{GHz})}\end{aligned}$$

The above equations presume a perfect match between noise-tee and its source and load impedance. As a result, a mismatch degrades the overall accuracy.

Using the theory on synthetic noise generation, as described in section 9.2.3 and 9.3.2, the value μ indicates that the noise bandwidth of the source was approximately $B \approx 1/(2\mu)^2 = 30.6 \text{ GHz}$.

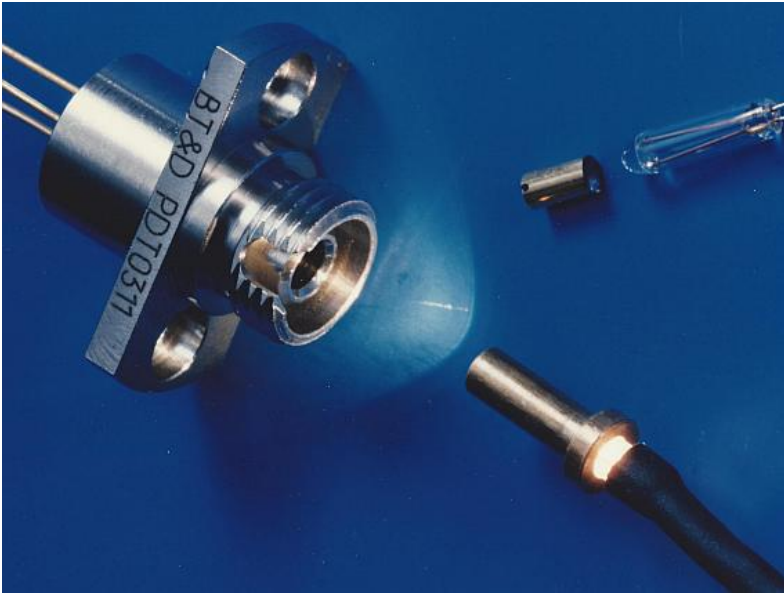


Fig 8.7 Illumination of a PIN photo-diode for shot-noise calibration. The light of a miniature lensed incandescent lamp is focused using a graded index lens, and inserted in the connectorized PIN photo-diode. See figure 8.8 for the associated setup of a shot-noise calibrator.

8.3.3. Calibration of synthetic noise, with shot-noise

The calibrated noise source of the calibration in subsection 8.3.2. is usually a thermal noise source or a calibrated noise source, initially calibrated with thermal noise. Shot-noise provides an attractive alternative.

The main advantage of our shot-noise calibrator is the high inherent accuracy when initial calibration with a primary noise standard is omitted. We observed corresponding results ($\approx 1\text{dB}$ difference) between the method described in section 8.3.2 and this section

8.3.3. Further the construction of the proposed shot-noise calibrator is simple, compared to thermal noise sources. The shot-noise level is reconstructed from dc-current measurements, and slightly adjusted using an initial calibration. The ambient temperature and illumination level are irrelevant.

As early as 1943, Breazeale, Beers, Waltz and Kuper applied shot-noise calibration methods using temperature-limited vacuum diodes [203: page 273-278]. The dc current flowing through an illuminated photo-diode provides another known method to generate shot-noise. This approach is known from RIN-measurements, usually using a tungsten lamp [835]. We developed a simple shot-noise calibration setup [801,802] for synthetic noise, using a (reverse biased) PIN photo-diode illuminated with an incandescent lamp. The measured photo current provides the required calibration level.

Basic principle

The random fluctuation associated with dc photo current provides the desired shot-noise. When all the electrons cross the PIN-barrier independently and fully at random, then pure shot-noise will be provided. From this, the following prior conditions are concluded for the generation of *pure* shot-noise:

- Any additional random modulation of the illumination intensity must be avoided because it will exceed the spectral density of the photo current above its shot-noise level. As a result, the illumination must originate from a 'clean' light source. We assume that incandescent lamps are sufficiently 'clean' compared to other inaccuracy aspects. Other light sources, such as semiconductor lasers (and to some extent LED's) are not applicable, since their (unknown) RIN-levels (relative intensity noise) cannot be ignored unless it is proved. Some alternatives with balanced detectors are discussed at the end of this subsection.
- Any correlation between the moments that the electrons cross the barrier will prevent the noise from being pure shot-noise. This is a well-understood mechanism in space-charge limited vacuum diodes [720, p15], and quantified by a noise suppression factor Γ . For PIN photo-diodes a value of $\Gamma=1$ is commonly accepted, however, lower values [802] are observed (approximately 10% or 1 dB). To our opinion this is *not* the result of measurement errors (assumed to be better than 0.3 dB). Proving our measurement accuracy is an aspect of further interest. As a result, the intensity spectrum of the shot-noise current in the PIN photo-diode, originating from an incandescent lamp, equals: $S_i = 2q \cdot I_{d0} \cdot \Gamma^2$, with $\Gamma \leq 1$.

Figure 8.8 shows the proposed shot-noise calibrator⁴, shunted to a second PIN photo-diode at the input of a low noise lightwave receiver. Using two diodes simplifies the construction. One diode construction is optimized for receiving synthetic noise via an optical fiber, and the other is optimized for illumination with incandescent lamps. The construction with the lamp is shown in the photograph of figure 8.7.

We developed a dedicated receiver with capacitive current-current feedback (see chapter 6 or [608]) to meet the high performance requirements on noise and linearity, without loss of available bandwidth. Illumination with the lamp resulted in a dc photo current of approximately $I_{d0} \approx 45 \mu A$, which yields approximately $\sqrt{S_i} \approx 3.8 pA/\sqrt{Hz}$ shot-noise. For comparison, the thermal noise current of a 50Ω resistor at room temperature is

⁴ The shot noise was generated with a very small incandescent lamp (lensed, 100mW), which fits in the hole of an FC/PC connectorized photo diode. The diode of the experimental setup was an InGaAs/Inp planar PIN photo diode, (BT&D PDT0311: 1.2pF, 0.85A/W typical).

18pA/ $\sqrt{\text{Hz}}$. This illustrates how weak practical shot-noise sources are. The thermal noise current of a 1.1k Ω resistor is approximately 3.8pA/ $\sqrt{\text{Hz}}$. As a result, the proposed shot-noise calibrator must be integrated with a low noise lightwave receiver, such as receivers with capacitive current-current feedback, as described in chapter 6.

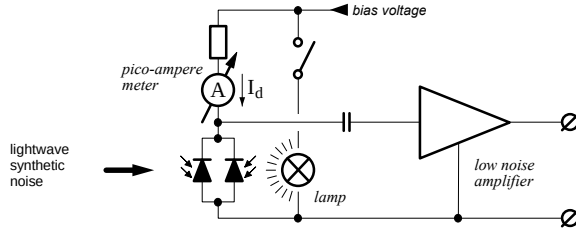


Fig 8.8 Circuit diagram of a shot-noise calibrator. The lamp illuminates a photo-diode and the resulting photo current is associated with shot-noise.

Calibration procedure

The shot-noise calibration is performed in two steps, similar to the noise calibration with a calibrated noise source and a noise-tee.

- In the first calibration state, the lamp is switched on while the lightwave synthetic noise generator is switched off. The resulting photo current is marked in this text as *calibration current* $I_{d,cal}$. The associated *calibration level* of the shot-noise is reconstructed from this dc current by calculation using $S_i = 2q \cdot I_{d0} \cdot \Gamma^2$.
- In the second calibration state, the lamp is switched off while the lightwave synthetic noise generator is switched on. The synthetic noise level is adjusted by an optical attenuator (see figure 8.2) to make the output noise level equal to the noise level of the first calibration state. As a result the synthetic noise level has become equal to the shot-noise level since both calibration states suffer from identical system noise contribution of the measurement setup. The resulting photo current is marked in this text as *reference current* $I_{d,ref}$.

Figure 8.9a shows the measured output spectrum of the shot-noise calibrator. Since the receiver noise is uncorrelated with the shot-noise and synthetic noise, and the calibration level is known, reconstruction of the individual input spectra is feasible. Figure 8.9b shows the result, using the extraction algorithms of section 8.4.

The (spectral) noise-current ratio μ of the synthetic noise is the ratio of the reference level $\sqrt{S_{i,ref}}$ and the reference current $I_{d,ref}$ and has the following value:

$$\mu \stackrel{\text{def}}{=} \frac{\sqrt{S_{i,ref}}}{I_{d,ref}} = \frac{\sqrt{2q \cdot I_{d,cal} \cdot \Gamma^2}}{I_{d,ref}} = \Gamma \cdot \frac{\sqrt{2q \cdot I_{d,cal}}}{I_{d,ref}}$$

The advantage of shot-noise calibration is that the construction is simple. The dc-current meter is the only device that requires an absolute accuracy.

The assumption of $\Gamma=1$ provides fair results, however, values of $\Gamma \approx 0.9$ are observed in the experimental setup. As a result, the noise level is 1 dB lower then was expected from pure shot-noise assumption. These are aspects of further investigations.

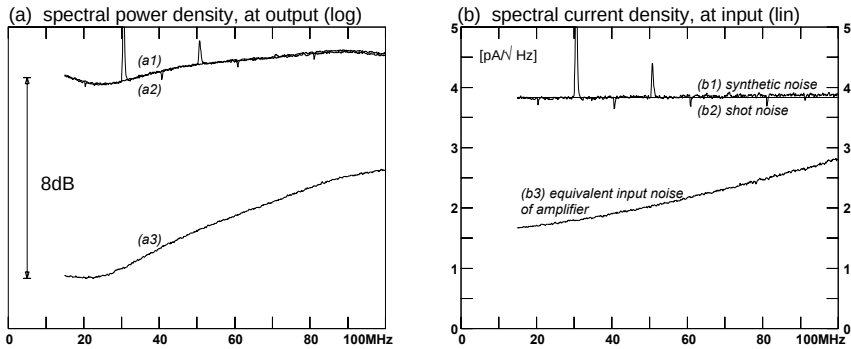


Fig 8.9 Measured output spectra (a) and reconstructed input spectra (b) of the shot-noise calibrator. The output spectra include the receiver noise and originate from (a1) lamp off, synthetic noise on, (a2) lamp on, synthetic noise off, and (a3) lamp off and synthetic noise off. The individual input spectra are unwrapped in figure b. The spikes originate from parasitic IM modulation in the synthetic noise source.

Illumination limitations

The excess-noise ratio (lamp on and lamp off) of the experimental shot-noise calibrator is relatively low. This is because shot-noise generators are in general weak noise sources. It is desired to increase the illumination, however this process suffers from principal limitations. This is summarized below:

- At first a photo-diode with low capacitance is required, to facilitate noise sources that are white over a sufficiently wide frequency band. Moreover, a low capacitance is crucial to facilitate low-noise amplification. Doubling the active area yields roughly doubling of capacitance and photo current, however the shot-noise increases with a factor $\sqrt{2}$. As a result, the active area of the PIN photo-diode must be chosen relatively small.
- Secondly the illumination originates from a (nearly) black-body radiator (hot filament of the incandescent lamp). This means that most of the optical power will not hit the active area of the photo-diode when projection methods with optical lenses⁵ are used. An optimal construction generates the desired photocurrent at minimal electrical lamp power, when the lamp temperature T is pre-defined. The best that can be achieved is that all black-body radiation of an area is focused (with perfect optics) on an area A with equal dimensions. In practice, the power, collected over area A , is significantly lower due to imperfect lenses (see construction in figure 8.7). Moreover, the power radiated over area A is lower when the filament wire is thinner than the area diameter.

Light sources with higher power commonly use longer filaments. This means that the additional power will not hit the active area, and that light sources with more

⁵ It is possible that alternative projection methods based on integrating spheres yield better results. These alternatives are not investigated.

optical power may not necessary improve the overall result. Improvement of the optical coupling between lamp and diode, and increase of filament temperature (halogen lamp) is more effective.

- Thirdly, the responsivity of the photo-diode is wavelength limited. Since the spectrum of a black body radiator is usually wider, no more than a small part of the collected power is used for photo current generation.

The radiation law of Stefan-Boltzmann provides an expression for the maximum collected power P within an area A . The radiation law of Planck provides the spectral components $p(\lambda)$ of this collected power. This *maximum* is quantified by:

$$P = \int_0^\infty p(\lambda) \cdot d\lambda = A \cdot \int_0^\infty \frac{(2\pi hc^2)/(\lambda^5)}{\exp((hc)/(\lambda kT)) - 1} \cdot d\lambda = A \cdot \frac{2\pi^5 k^4}{15h^3 c^2} \cdot T^4 = A \cdot \sigma \cdot T^4$$

$$h = 6.626176 \cdot 10^{-34} \text{ [J}\cdot\text{s]}$$

Planck constant

$$c = 2.99792458 \cdot 10^8 \text{ [m/s]}$$

Speed of light in vacuum

$$k = 1.380662 \cdot 10^{-23} \text{ [J/K]}$$

Boltzmann constant

$$\sigma = 5.67032 \cdot 10^{-8} \text{ [W/(m}^2\cdot\text{K}^4)]$$

Stefan-Boltzmann constant

For example, consider a PIN photo-diode with $d=70 \mu\text{m}$ diameter, an active area of $A = \pi/4 \cdot d^2$, and a filament with temperature $T=2500 \text{ K}$. Then the collected illumination will not exceed the power level of $P=8.5 \text{ mW}$. Furthermore assume a PIN photo-diode with a wavelength window of 1000 nm and 1600 nm , then approximately 32% of the collected power is used for photo current generation. This yields no more than 2 mA photo-current for a diode with 0.75 A/W sensitivity.

The maximum photo current, generated in our experimental setup with a miniature lensed lamp and an additional graded-index lens, was approximately 0.3 mA . This demonstrates that the shot-noise generator in the experimental setup is not optimal, however huge improvements are essentially impossible.

Alternative shot-noise sources

The unknown RIN (relative intensity noise) of LED's and lasers prevent them from direct use as shot-noise source in a shot-noise calibrator. The use of balanced receivers may facilitate an adequate suppression of the RIN of lasers [e.g. Kaspar et al: 805] and of LED's [e.g. Machida and Yamamoto: 806].

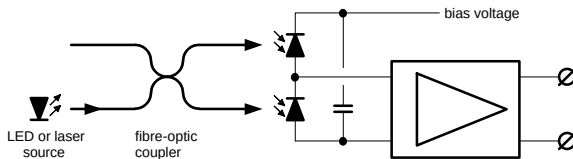


Fig 8.10 Balanced configuration to generate shot-noise with lightwave sources that suffer from RIN.

Figure 8.10 shows the basic configuration. The light beam of an LED or a laser source is coupled to a fiber and equally split in two beams. Each beam illuminates a photo-diode, and generates both shot-noise and RIN. Both magnitude and phase of the two branches are adjusted to perform perfect balance: equal beam splitting factor, equal fiber length and equal photo-diodes.

Since the RIN originates from the same light source, the associated noise currents in both diodes are fully correlated. As a result, the RIN is perfectly suppressed by subtraction. Since the shot-noise originates in the PIN photo-diodes, the associated noise currents in both diodes are fully uncorrelated. As a result, the intensity spectra of both shot-noise current sources increase by addition.

The balanced configuration may be used as an alternative for shot-noise generation with an incandescent lamp. However, it is not further discussed in this text.

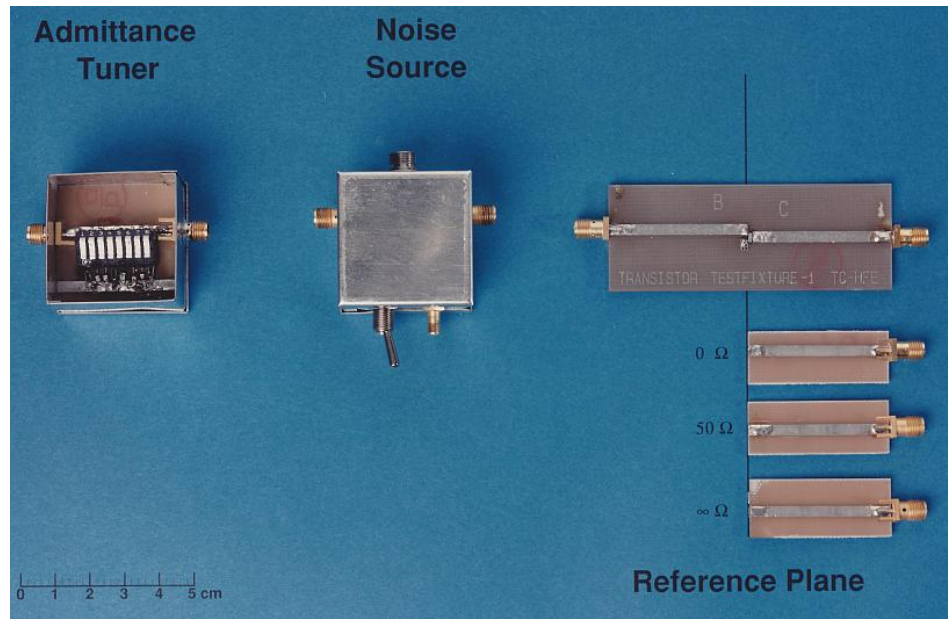


Fig 8.11 Practical setup for measuring noise, relative to an (arbitrary located) reference plane. In this example, a transistor is the amplifier under test, as illustrated in figure 8.12.

8.3.4. Transformation of calibrated noise to arbitrary reference planes

One of the most startling features of the noise-tee is the variation of output impedance with a simple one-port impedance tuner. When this impedance is known from impedance measurements, and shunted with a current source with white noise spectrum, then the spectrum of the noise voltage is easily extracted. However, the location of the output reference plane of the noise-tee does not coincide with the noise-tee photo-diode, which makes the intensity spectrum of the output noise subject to frequency variations. In the restricted case that one side of the noise-tee is matched to the characteristic impedance Z_0 of the noise-tee, the synthetic noise spectrum is preserved at the noise-tee output (see figure 8.2). This makes the calibration of the previous subsections 8.3.2 and 8.3.3 directly applicable for noise measurements with fixed source impedances over a wide frequency range.

The general case of arbitrary reference planes requires an additional transformation of the noise from the diode reference plane to the chosen reference plane, by using the (measured) two-port parameters of the enclosed two-port. This subsection describes the required mathematics to perform this transformation.

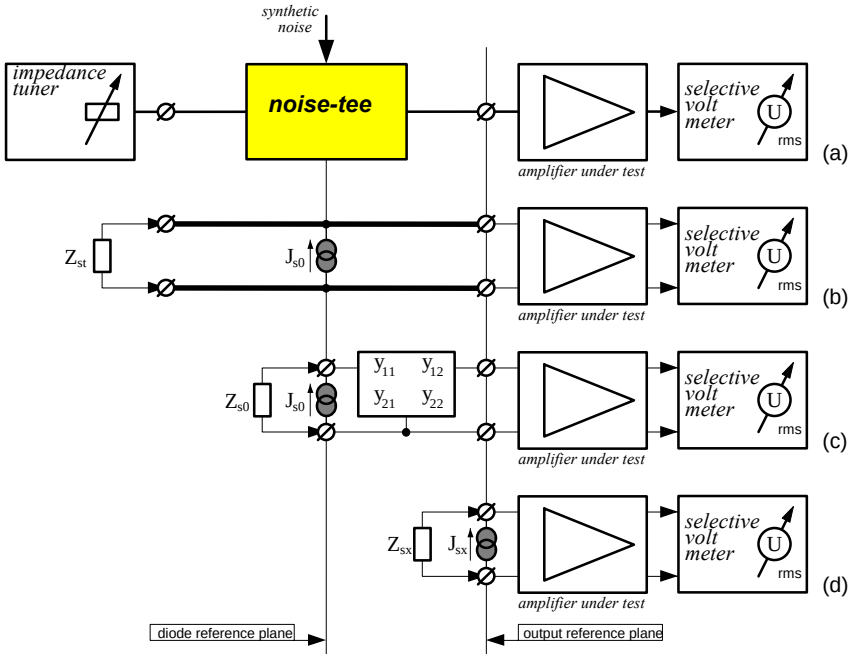


Fig 8.12 Mis-matched noise-tee configuration for noise measurements on amplifiers at specified source impedance.

Figure 8.12a shows a mis-matched noise-tee configuration to perform noise measurements on amplifiers at specified source impedance Z_{sx} . These measurements are of interest when the application of the amplifier under test is intended for sources with impedance Z_{sx} . The impedance tuner is adjusted to imitate that source impedance Z_{sx} of interest (observed at the output reference plane of the noise-tee).

Figure 8.12b shows the interior transmission line of the noise-tee that links the tuner impedance $Z_{st}(\omega)$ with the input of the amplifier under test. The (white) synthetic noise is represented by the (complex) noise current spectrum⁶ $J_{s0}(\omega)$. The (real) intensity spectrum $S_{Js}(\omega)$ of current $J_{s0}(\omega)$ is extracted by one of the previous discussed calibration methods (see section 7.1 for definitions).

Figure 8.12c shows the equivalent noise current spectrum $J_{sx}(\omega)$ that is the combination of synthetic noise and thermal noise, observed from the diode reference plane. The interior transmission line transforms tuner impedance Z_{st} into impedance Z_{s0} . Figure 8.12d shows the equivalent impedance and noise $Z_{sx}(\omega)$ and $J_{sx}(\omega)$, observed from the output reference plane. These values are to be extracted.

Assume that all two-port parameters of the right half of the noise-tee are known, for instance measured in s-parameter format. Transform them into y-parameter format for calculation convenience. Additionally assume the output impedance Z_{sz} is also known, e.g. by measurement. Using the general embedding and de-embedding equations of section 2.3.1, the transformations between J_{s0} and J_{sx} become:

$$\left. \begin{aligned} J_{sx} &= \left(\frac{-y_{21}}{Y_{s0} + y_{11}} \right) J_{s0} & Y_{sx} &= 1/Z_{sx} = \left(\frac{\Delta_y + Y_{s0} \cdot y_{22}}{Y_{s0} + y_{11}} \right) \\ J_{s0} &= \left(\frac{y_{12}}{Y_{sx} - y_{22}} \right) J_{sx} & Y_{s0} &= 1/Z_{s0} = \left(\frac{\Delta_y - Y_{sx} \cdot y_{11}}{Y_{sx} - y_{22}} \right) \end{aligned} \right\} \Rightarrow J_{sx} = \left(\frac{Y_{sx} - y_{22}}{y_{12}} \right) \cdot J_{s0}$$

The total hot level of the output noise is the composition of thermal noise (cold noise), due to tuner impedance and noise-tee losses, and the noise current spectrum J_{sx} (excess-noise). In a previous chapter, it is discussed that the intensity spectrum of a thermal noise current is easily extracted from admittance measurements. Since this cold noise current and excess-noise current are uncorrelated, and the (spectral) noise current ratio μ of the synthetic noise source is known, the various intensity spectra become:

$$\begin{aligned} S_{Js0,excess} &= (\mu \cdot I_d)^2 && \text{originating from synthetic noise source} \\ S_{Js0,cold} &= 4 \cdot kT \cdot \text{real}(Y_{s0}) && \text{originating from thermal noise} \\ S_{Js0,hot} &= 4 \cdot kT \cdot \text{real}(Y_{s0}) + (\mu \cdot I_d)^2 && \text{total hot noise level at output reference plane} \end{aligned}$$

$S_{Js0,hot} = (\mu \cdot I_d)^2 + 4 \cdot kT \cdot \text{real} \left\{ \frac{\Delta_y - Y_{sx} \cdot y_{11}}{Y_{sx} - y_{22}} \right\}$	<i>observed at diode reference plane</i>
$S_{Jsx,hot} = (\mu \cdot I_d)^2 \cdot \left \frac{Y_{sx} - y_{22}}{y_{12}} \right ^2 + 4 \cdot kT \cdot \text{real}(Y_{sx})$	<i>observed at output reference plane</i>

When both the diode reference plane and the output reference plane are (externally) accessible for measurement, then the transforming two-port parameters are extracted from two-port measurements. If not, the two-port parameters of the right half of the noise-tee must be made available from mathematical halving of the complete two-port. This is discussed in the succeeding subsection 8.3.5.

⁶ Complex noise spectra $J(\omega)$ are used only in intermediate results for calculation convenience, because they comprise amplitude as well as phase information. Since noisy signals are not deterministic by definition, phase information is not available, and amplitude information is restricted to a mean value in a small frequency band. Measurements are restricted to spectral power densities $S_J(\omega) = \langle |J(\omega)|^2 \rangle_{\Delta\omega}$ of noisy signals.

8.3.5. Mathematical halving of the noise-tee

The two-port parameters of the internal sections of the noise-tee are not available by direct measurement, because the diode reference plane in figure 8.12 is inaccessible. Nevertheless, the symmetric two-port construction of the noise-tee facilitates another approach.

The noise-tee is a cascade of two identical but reversed two-ports, separated by the diode reference plane. When the overall cascade is measured, the internal sections may be extracted by a mathematical operation, similar to the 'root' of a matrix. An exact solution is essentially impossible, because a unique solution does not exist⁷. However, the most plausible solution is feasible in the case that an interior circuit model is available. The symmetrical construction of figure 8.1 and 8.14 is proposed to facilitate mathematical 'halving' of the noise-tee.

Figure 8.13 shows an equivalent circuit model of the internal noise-tee: a cascade of two transmission lines, that encloses a shunt impedance Z in the center. Due to imperfect balancing, the two transmission lines have different lengths. The shunt impedance represents the imperfect compensation of the diode capacitance by the micro strip slit (see figure 8.1).

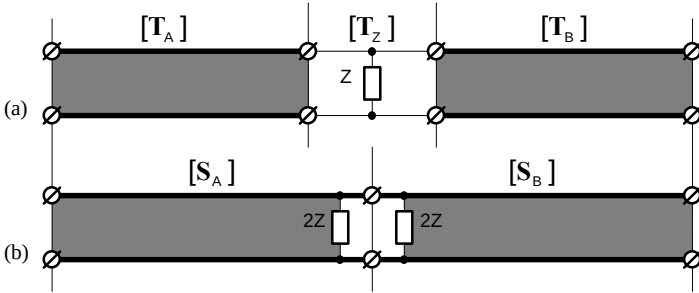


Fig 8.13 Equivalent circuit model of the noise-tee, to extract the halved two-ports from overall two-port measurements. In (a) the noise-tee is divided into three sections, while in (b) the shunt impedance Z is combined with the transmission line sections.

Figure 8.13a shows the cascade of idealized two-ports, using s-parameters rearranged in a transmission matrix. Let $x \triangleq \frac{1}{2} \cdot (Z_0/Z)$ be a shortcut to represent the shunt impedance, and let s_a and s_b represent the transmission of waves through the transmission lines. Then it can be demonstrated that the cascaded two-port has the following T-matrix and S-matrix form (see section 2.3.1 and appendix A):

$$\mathbf{T} = \mathbf{T}_A \cdot \mathbf{T}_Z \cdot \mathbf{T}_B = \begin{bmatrix} 1/s_a & 0 \\ 0 & s_a \end{bmatrix} \cdot \begin{bmatrix} 1-x & -x \\ x & 1+x \end{bmatrix} \cdot \begin{bmatrix} 1/s_b & 0 \\ 0 & s_b \end{bmatrix} = \begin{bmatrix} (s_a \cdot s_b) \cdot (1-x) & -(x) \cdot (s_b/s_a) \\ (s_a/s_b) \cdot (x) & (1+x) \cdot (s_a \cdot s_b) \end{bmatrix}$$

⁷ The noise-tee is a symmetrical and reciprocal two-port, and is characterized by two unique complex numbers ($s_{11}=s_{22}$, and $s_{12}=s_{21}$). The mathematical halving must provide three unique complex numbers to represent its non-symmetrical halves. As a result, the set of two equations is incomplete to find a unique solution for three unknown.

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} = \frac{1}{t_{11}} \cdot \begin{bmatrix} t_{21} & \Delta_t \\ 1 & -t_{12} \end{bmatrix} = \frac{1}{(1-x)} \cdot \begin{bmatrix} (s_a \cdot s_b) \cdot x & (s_a \cdot s_b) \\ (s_a \cdot s_b) & x \cdot (s_a \cdot s_b) \end{bmatrix}$$

The s-parameters of the (reciprocal) two-port are made available from measurements, which result in numerical values for x , s_a and s_b . Figure 8.13b shows how the shunt impedance is combined with the transmission line sections. The s-parameters of the scattering matrices $\mathbf{S}_A$ and $\mathbf{S}_B$ represent the two noise-tee sections:

$$\mathbf{S}_A = \frac{1}{(1-x/2)} \cdot \begin{bmatrix} (s_a^2) \cdot (x/2) & s_a \\ s_a & (x/2) \end{bmatrix}$$

$$\mathbf{S}_B = \frac{1}{(1-x/2)} \cdot \begin{bmatrix} (x/2) & s_b \\ s_b & (x/2) \cdot (s_b^2) \end{bmatrix}$$

using $\left\{ \begin{array}{l} \delta \stackrel{\text{def}}{=} (s_b/s_a) = \sqrt{(s_{22}/s_{11})} \approx 1 \\ x = \delta \cdot (s_{11}/s_{12}) \\ s_a = \sqrt{(s_{12} \cdot (1-x)/\delta)} = s_b/\delta \\ s_b = \sqrt{(s_{12} \cdot (1-x)\delta)} = s_a/\delta \end{array} \right\}$

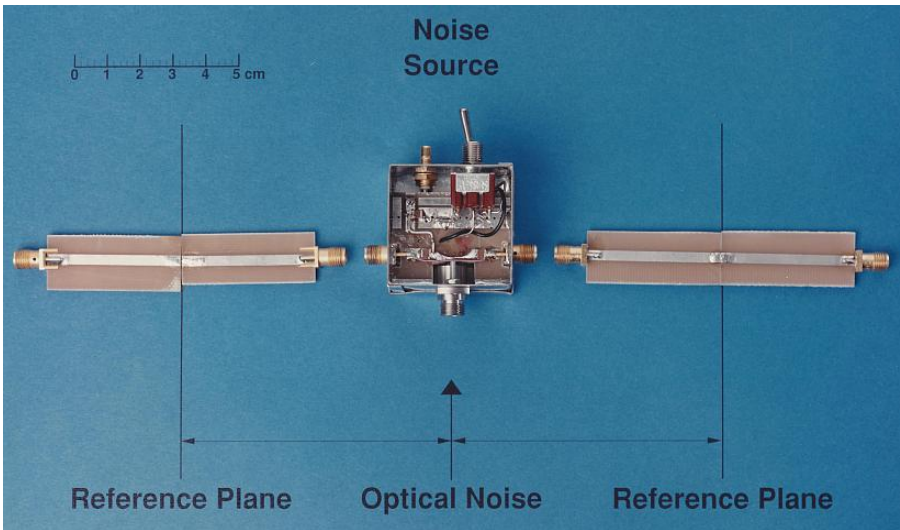


Fig 8.14 Mathematical halving of the lightwave noise-tee. The noise-tee is measured as electrical two-port, relative to external reference planes.

Practical aspects

The value δ represents the unbalance between the two sections. Its value is extracted from reflection measurements (s_{11} and s_{22}), however in well-designed noise-tees this

reflection is close to zero. As a result, the proposed extraction method for δ will yield unreliable results in case of low reflection values.

A fair solution is using an estimation for δ , e.g. by setting $\delta=1$, based on the mechanical unbalance in the construction. More reliable values are obtained for δ from unbalance measurements using the photo-diode of the noise-tee (see figure 8.15). Modulate the illumination with a periodic signal, and detect the signal unbalance between the two output ports of the noise-tee. The (complex) ratio between both signals provides δ at the modulated frequency.

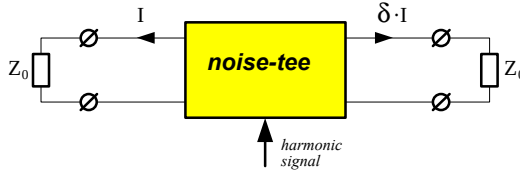


Fig 8.15 Noise-tee unbalance measurement using the lightwave input. The ratio δ between both output currents is a measure for the unbalance between the two halves.

Another practical aspect is the sign uncertainty in the extraction of s_a and s_b . The root function has two solutions, and this may result in a sign uncertainty when wavelength is shorter than the noise-tee length. This is illustrated in figure 8.16. A reliable solution is obtained from phase unwrapping of s_a , using succeeding frequencies. The phase of s_a , as function of the frequency, has a saw-tooth shape. Add a multiple of the value π to this phase when it wraps, to make it a continuous function. The most reliable unwrapping is obtained when this decision is based on a transmission line model of the noise-tee, using line length and frequency information.

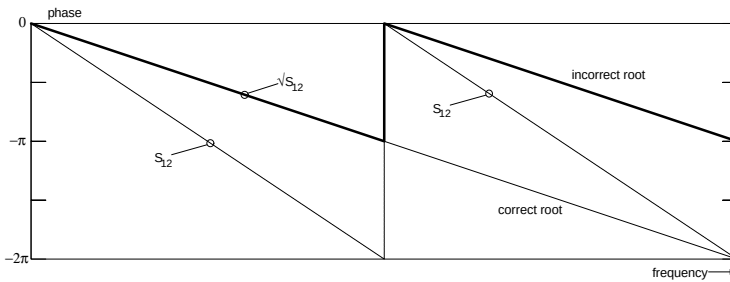


Fig 8.16 Sign uncertainty in the extraction of s_a and s_b . When the phase of s_{12} wraps from -2π to 0 then the phase of the root function $\sqrt{s_{12}}$ will wrap from π to 0. As a result, the sign of $\sqrt{s_{12}}$ swaps and yields an illegal value for s_a and s_b .

The extracted internal shunt impedance Z is indicative for the compensation performance of the diode capacitance by the micro strip slit (see figure 8.1). In the experimental noise-tee, the diode capacitance was reduced from 1 pF to below 0.3 pF, simply by removal of an empirically determined amount of copper. This demonstrates how effective the proposed compensation is.

The experimental noise-tee is based on a $50\ \Omega$ microstrip line, 3 mm wide and 38 mm long. The substrate is a standard glass-epoxy printed circuit board, 1.5 mm thick, 25 mm wide and 38 mm long.

8.3.6. Conclusions

Two convenient calibration methods for lightwave synthetic noise are demonstrated, using a ($50\ \Omega$) calibrated noise source and using shot-noise. Although the noise of the demonstrated shot-noise calibrator is significantly lower than thermal noise of $50\ \Omega$ resistors, the 'hot' shot-noise level is 8 dB higher than the 'cold' noise level. Furthermore, theoretical limitations are discussed of the maximum shot-noise level that incandescent lamps can generate in the proposed shot-noise calibrator.

The validation of the proposed calibration methods is restricted to reference planes that coincide with the PIN photo-diode in the noise-tee, unless both ports of the noise-tee are matched to its characteristic impedance. We demonstrated how to transform this calibration to arbitrary reference planes.

8.4. Noise level measurements using matched sources

The previous sections described various white noise sources and various methods to calibrate their noise level. This subsection applies these techniques to ratio noise measurements. A novel measurement method is described to measure the input noise of lightwave receivers, observed from an *optical* reference plane [801, 803]. To demonstrate its superior performance, our measurement method are compared with various alternative methods.

Ratio noise measurements with two calibrated noise levels (hot and cold) are commonly used for measurements on electrical receivers. Hence this section 8.4 is focused on multi-level noise measurements, applied to lightwave receivers.

This section describes a novel noise-extraction algorithm that takes the benefits of *multi-level* noise sources in ratio measurements. The proposed noise extraction algorithm minimizes the *relative*⁸ error in the reconstructed equivalent input noise levels. This provides the most plausible solution from detected output noise levels with comparative detection errors.

The section is restricted to noise sources, matched to the transmission line that links the source with the device under test. This restriction makes the noise position-*invariant*, which means that a location shift of the reference plane will not result in a noticeable chance of observed reference noise levels.

For electrical noise sources, this condition implies that the output impedance of the source is matched to the linking transmission line. As a result, the observed noise level as well as the source impedance are position invariant, observed from any reference plane between source and DUT (device under test).

For lightwave noise sources, this restriction implies that reflected waves are completely absorbed in the source. Furthermore it implies that the isolation between laser and fiber is adequate to prevent the laser from noticeable optical feedback effects.

The restriction of this section simplifies the measurement significantly. It makes complicated reference plane transformations, as described in section 8.3.4, superfluous. The more general case, with arbitrary mis-matched source impedances, is discussed in the succeeding section 8.5.

8.4.1. Basic principles and definitions

Figure 8.17 shows a block diagram of a ratio measurement setup. It starts with a (calibrated) variable noise source, generating (white) noise. The (single sided) intensity spectrum of the output noise level is set to reference input level S_r . The known noise level of the variable noise source as well as the unknown noise level of the device under test is amplified and detected with a frequency selective volt meter. The unknown system noise increases the detected voltage level.

⁸ The algorithms that we have found in the literature focus on the minimization of the absolute error (see section 8.5.1 on the state of the art). We assume that such an approach result in sub-optimal solutions, and that minimization of the relative error gives better performance.

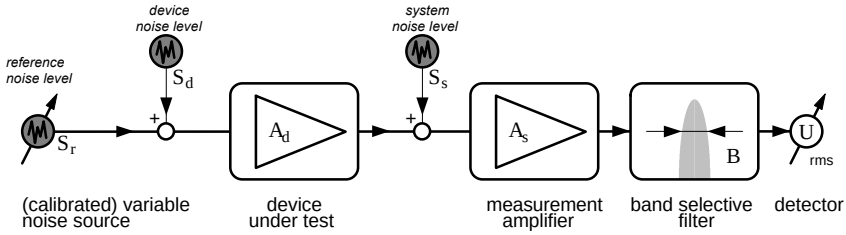


Fig 8.17 Block diagram of a ratio noise measurement setup. The device noise level is measured using calibrated noise. The gain of the device under test is unknown, and all sources are assumed to be uncorrelated.

S_s = intensity spectrum of system equivalent input noise

S_d = intensity spectrum of device equivalent input noise

S_r = intensity spectrum of reference input noise

The detector detects the rms-voltage of the noise that has passed a selective band filter. It detects the values:

U_s when the device under test is switched off,

U_d when the device is activated but the variable noise source is switched off, and

U_r when the noise source as well as the device are activated,

All noise sources are assumed to be uncorrelated, which means that two noise currents add on a power base. As a result, the detected voltages are quantified by:

$$U_s \stackrel{\text{def}}{=} \sqrt{A_s^2 \cdot B \cdot S_s} \quad = \text{detected system noise}$$

$$U_d \stackrel{\text{def}}{=} \sqrt{A_s^2 \cdot B \cdot S_s + (A_d \cdot A_s)^2 \cdot B \cdot S_d} \quad = \text{detected device noise}$$

$$U_r \stackrel{\text{def}}{=} \sqrt{A_s^2 \cdot B \cdot S_s + (A_d \cdot A_s)^2 \cdot B \cdot S_d + (A_d \cdot A_s)^2 \cdot B \cdot S_r} \quad = \text{detected reference noise}$$

The device noise level S_d is extracted from the measured voltage levels (U_s , U_d , U_r) and the reference noise level (S_r) by solving the above equations. When B and A_s are known, then the extraction process provides the device gain A_d too. To simplify these equations, we define:

$$G \stackrel{\text{def}}{=} (A_d \cdot A_s \cdot \sqrt{B})^2 \quad \text{virtual power 'gain' of measurement setup}$$

$$S_{ds} \stackrel{\text{def}}{=} (S_d + S_s / A_d^2) \quad \text{intensity spectrum of equivalent input noise}$$

$$P_s \stackrel{\text{def}}{=} U_s^2 = G \cdot (S_s / A_d^2)$$

$$P_d \stackrel{\text{def}}{=} U_d^2 = G \cdot (S_s / A_d^2 + S_d) = G \cdot (S_{ds})$$

$$P_r \stackrel{\text{def}}{=} U_r^2 = G \cdot (S_s / A_d^2 + S_d + S_r) = G \cdot (S_{ds} + S_r)$$

The measurement of voltage level U_d requires the variable noise source to be switched between 'on' and 'off'. When the device under test is a lightwave receiver, and thus the variable noise source must provide *optical* noise, then a perfect 'off' state exists. However, when the variable noise source must provide *electrical* noise, then the thermal noise of the noise source impedance limits the minimum reference noise level.

Simply removal of the noise source will affect the device noise level S_d and yields unpredictable⁹ results. This demonstrates that it is not self-evident that the value U_d is available for detection. As a result, when U_d is lacking then the extraction of the device noise level S_d requires at least two or more measurements of U_p , associated with different reference levels.

In the following sections, the required extraction algorithms will be discussed in more detail.

8.4.2. Extraction of input noise level using hot/cold noise sources

As early as 1953, IRE standards recommended [817,818,821] ratio noise measurements using two calibrated noise levels. The recommended method requires a variable noise source, of which the output noise level can be switched between 'high' and 'low' or rather between 'hot' and 'cold'. The restricted case of two reference levels (S_{r1} , S_{r2}), enables the use of a simple algorithm to extract the device spectrum S_d from the measurements (U_s , U_{r1} , U_{r2}). It can be demonstrated that:

$$\begin{aligned} G &= \frac{S_{r2} - S_{r1}}{P_{r2} - P_{r1}} \\ S_{ds} &= \frac{P_{r1} \cdot S_{r2} - P_{r2} \cdot S_{r1}}{P_{r2} - P_{r1}} \\ S_d &= \frac{(P_{r1} - P_s) \cdot S_{r2} - (P_{r2} - P_s) \cdot S_{r1}}{P_{r2} - P_{r1}} \end{aligned}$$

see section 8.4.1
for the definition
of G , S and P .

The higher the ratio (S_{r2}/S_{r1}) is, the more precise S_d is extracted, provided that the device under test is not overloaded. Commonly specified values for the *excess-noise ratio*¹⁰ (see section 8.1) of calibrated 50 Ω noise sources ranges from 5 dB to 20 dB. This means that (S_{r2}/S_{r1}) ranges from 2^2 to 10^2 . The 'cold' noise level is commonly the thermal noise of a 50 Ω resistor at the standard reference temperature $T=290$ K.

This method is often referred to as the Y-factor method, probably due to the ratio $Y=P_{r2}/P_{r1}$ as used in formula 10 of [821].

8.4.3. Extraction of input noise level using multi-level noise sources

The choice of noise measurements with two noise levels is dictated by the limitations of commonly used electrical noise sources. The lightwave noise-tee, proposed in section 8.2, is much more flexible. Continuous variation of noise level is easy, without loss of accuracy or noticeable change of source impedance. This facilitates the use of more than one calibrated hot and cold noise level.

Multi-level noise measurements are attractive to reduce measurement errors by redundancy. Measurement errors are unavoidable, for instance quantization errors and distortion errors. Using more than two measurement levels, spread out over the dynamic

⁹ While measuring equivalent input noise, the noise source impedance must correspond with the signal source impedance for which the amplifier under test will be used. If not, the measured device noise will not correspond with the device noise under normal operation.

¹⁰ The *excess noise ratio* is defined (in dB) as $ENR = 10 \cdot \log(S_{r2}/S_{r1} - 1)$

range of interest, enables the extraction of a more plausible solution. That solution should spread errors out over all measurements, avoiding specific errors being improbably higher than all other errors. This approach may improve the overall measurement accuracy.

Conventional hot/cold measurements provide sets of equations having unique solutions. This does not hold for an over-determined set of equations, and the device noise extraction algorithm should select the most plausible solution.

A robust algorithm has been resulted from this study, that extracts S_d from multi-level measurements, optimized in a least-squares sense. Let the column vector $\mathbf{P}_r = [U_{r1}^2, U_{r2}^2, \dots, U_{rm}^2]^T$ represent the detected squared voltages U_r associated with the reference noise levels $\mathbf{S}_r = [S_{r1}, S_{r2}, \dots, S_{rm}]^T$. When more than two levels are measured then there will be no exact solution obtained for G and S_{ds} from the matrix equation: $\mathbf{P}_r = (S_{ds} + \mathbf{S}_r) \cdot G$.

All individual measurements P_{rx} are associated with a small deviation from the actual value. As a result, all expressions listed below are adequate.

$$\begin{aligned} G \cdot (S_{ds} + S_{rx}) &\approx P_{rx} \\ G \cdot (S_{ds} + S_{rx}) &= P_{rx} + \partial P_x \\ G \cdot (S_{ds} + S_{rx}) &= P_{rx} \cdot (1 + \delta_x) \end{aligned} \quad \begin{aligned} \partial P_x &\text{ represents an absolute measurement error} \\ \delta_x &\text{ represents a relative measurement error} \end{aligned}$$

An arbitrary solution can be found simply, by minimizing an arbitrary error (e.g. ∂P) with proper mathematics, however this will not necessary result in an optimal solution.

Assume that the accuracy of the reference levels is superior, compared to the random measurement errors. Then, an optimal reconstruction of G and S_{ds} is associated with optimal *relative* accuracy for the reconstructed values $(S_{ds} + S_{rx})$. This yields optimal relative accuracy for $G \cdot (S_{ds} + S_{rx})$ which is equivalent with minimal (rms) average of the relative errors $\delta = [\delta_1, \delta_2, \dots, \delta_n]^T$. As a result, *equal* minimization of the errors ∂P_x will *not* yield the most plausible solution, while *equal* minimization of the errors δ_x will do.

Now, the following matrix equation must hold:

$$\begin{bmatrix} P_{r1} \cdot (1 + \delta_1) \\ P_{r2} \cdot (1 + \delta_2) \\ \dots \\ P_{rm} \cdot (1 + \delta_n) \end{bmatrix} = G \cdot S_{ds} + G \cdot \begin{bmatrix} S_{r1} \\ S_{r2} \\ \dots \\ S_{rm} \end{bmatrix}$$

$$\begin{bmatrix} 1 + \delta_1 \\ 1 + \delta_2 \\ \dots \\ 1 + \delta_n \end{bmatrix} = \begin{bmatrix} 1/P_{r1} & S_{r1}/P_{r1} \\ 1/P_{r2} & S_{r2}/P_{r2} \\ \dots & \dots \\ 1/P_{rm} & S_{rm}/P_{rm} \end{bmatrix} \cdot \begin{bmatrix} G \cdot S_{ds} \\ G \end{bmatrix} = \begin{bmatrix} 1/P_{r1} & (S_{r1}/S_0)/P_{r1} \\ 1/P_{r2} & (S_{r2}/S_0)/P_{r2} \\ \dots & \dots \\ 1/P_{rm} & (S_{rm}/S_0)/P_{rm} \end{bmatrix} \cdot \begin{bmatrix} G \cdot S_{ds} \\ G \cdot S_0 \end{bmatrix}$$

S_0 is an arbitrary scaling factor to reduce numerical roundoff errors.
Optimal scaling is performed when S_0 is set to a value close to an average reference value S_r .

This matrix equation has $(1 + \delta) = A \cdot x$ as general form. The most plausible solution vector x must minimize vector δ , for instance in a least-squares sense. Using the over-determined left-hand matrix division, as defined in appendix B, then a solution vector

$\mathbf{x}=\mathbf{A}\mathbf{1}$ fulfills the requirement that the inner product $(\delta \delta')$ is minimized. This yields for the most plausible solution for G and S_{ds} :

$$\begin{bmatrix} G \cdot S_{ds} \\ G \cdot S_0 \end{bmatrix} = \begin{bmatrix} 1/P_{r1} & (S_1/S_0)/P_{r1} \\ 1/P_{r2} & (S_2/S_0)/P_{r2} \\ \dots & \dots \\ 1/P_{rn} & (S_n/S_0)/P_{rn} \end{bmatrix} \setminus \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$$

See section 8.4.1 for the definition of G , S and P .
 S_0 is an arbitrary scaling factor.

The scaling factor S_0 does not affect the solution, however a proper chosen value avoids round-off errors due to finite numerical precision. A good choice for this scaling factor is an arbitrary value, close to an average value of the reference levels S_r .

Error weighing

In the above equations, all levels are equal weighted. This means that an error in some level A is as important as another error in some level B. For example, when 90% of the measurements are performed at nearly equal levels and the remaining 10% at another level, than the above algorithm will favor the majority of the (correlated) errors at the expense of the minority. This is because the average error reduces when these (correlated) errors improve.

In this example, this effect is undesirable when all measurements were performed with equal accuracy, since it wrongly favors some specific levels. On the other hand, when the highest level is associated with the most erroneous measurement (e.g. distortion), then this effect may be crucial. The optimal solution for this effect is *weighing* of individual errors. It will relax the impact of individual errors on the overall accuracy.

Let $\mathbf{w}=[w_1, w_2, \dots, w_n]$ be an appropriate set of weighing factors to define the desired minimization of $[w_1 \cdot (S_{ds} + S_{r1}), w_2 \cdot (S_{ds} + S_{r2}), \dots, w_n \cdot (S_{ds} + S_{rn})]$. Then this minimization is equivalent with minimization of $[w_1 \cdot \delta_1, w_2 \cdot \delta_2, \dots, w_n \cdot \delta_n]$. Using the same over-determined left-hand matrix division as discussed before, it can be demonstrated that:

$$\begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix} + \begin{bmatrix} w_1 \cdot \delta_1 \\ w_2 \cdot \delta_2 \\ \dots \\ w_n \cdot \delta_n \end{bmatrix} = \begin{bmatrix} w_1/P_{r1} & (w_1/P_{r1}) \cdot (S_1/S_0) \\ w_2/P_{r2} & (w_2/P_{r2}) \cdot (S_2/S_0) \\ \dots & \dots \\ w_n/P_{rn} & (w_n/P_{rn}) \cdot (S_n/S_0) \end{bmatrix} \cdot \begin{bmatrix} G \cdot S_{ds} \\ G \cdot S_0 \end{bmatrix}$$

$$\begin{bmatrix} G \cdot S_{ds} \\ G \cdot S_0 \end{bmatrix} = \begin{bmatrix} w_1/P_{r1} & (w_1/P_{r1}) \cdot (S_1/S_0) \\ w_2/P_{r2} & (w_2/P_{r2}) \cdot (S_2/S_0) \\ \dots & \dots \\ w_n/P_{rn} & (w_n/P_{rn}) \cdot (S_n/S_0) \end{bmatrix} \setminus \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix}$$

See section 8.4.1 for the definition of G , S and P .
 S_0 is an arbitrary scaling factor.
 w_x are weighing factors

Set all weighing factors to an equal value, e.g. $w=1$, when all reference levels should equally pronounce in the reconstruction process. All specific levels with limited precision must be weighted with a lower value. A specific weighing value of $w_x=0$ will exclude a specific level.

8.4.4. Measurement of lightwave receiver noise

The input noise of the front-end receiver of an optical communication system is an important parameter for calculating the overall system sensitivity. Many lightwave receivers are reported in the literature, that focused on sensitivity improvement. Specification of the equivalent noise current (in $\text{pA}/\sqrt{\text{Hz}}$) is commonly used, to enable comparison with other reported receivers, or to claim new sensitivity records.

State of the art

Investigating commonly used noise measurement methods, we found no more than a few publications that mentioned their method to measure the receiver noise. The following methods were found:

- Direct noise measurements are often mentioned (see section 8.1). They measure output noise (usually with a spectrum analyzer), measure receiver gain or estimate it from what it should be, and reconstruct from these quantities the equivalent input noise [807, 808, 809, 810, 811, 812, 813, 814, 815].
- Kaspar et al. [805] compared equivalent input noise of a *balanced* receiver with white noise using a laser as light source. The balanced configuration facilitates suppression of laser RIN and the remaining noise level is (close to) the shot-noise level. This ratio calibration method has been discussed in section 8.3.3. In fact, they did not work out the method completely, however they enabled the reader to reconstruct the noise level from the ratio of the reported output spectra.
- Machida and Yamamoto [806] measured squeezing of light using a similar ratio noise measurement. They generated white noise in a balanced pair of photo-diodes using an LED light source (see section 8.3.3).
- Incidentally, the noise was not measured but estimated from specified noise parameters of the transistors that were used.

Most reported noise measurements did not mention how this noise level was measured. We assume that direct noise measurements are commonly used since instruments that facilitates lightwave ratio noise measurements are currently not commercially available.

Reliability of commonly used measurements

Most of the investigated papers did not indicate the reliability of the published values. The above overview is assumed to be indicative for the reliability of reported noise values since direct noise measurements suffer from poor accuracy performance. Many high-performance spectrum analyzers are specified with $\pm 2\text{dB}$ uncertainty for harmonic signals. Additional errors, that are associated with noise measurements, originate from uncertainties in noise correction, mismatch between receiver and instrument, and gain uncertainties.

The simplest noise correction divides the measured output noise spectrum by the root of the measurement bandwidth. Additional corrections up to 2dB are required for various noise corrections [804], however none of the above direct noise measurements mentioned what effects are taken into account, how these correction values were obtained and with what accuracy. This is relevant information since spectrum analyzers are not always equipped with specified noise markers.

Furthermore, we observed that it is no common use to specify *what* is measured. The information is usually restricted to the words *noise power spectrum* or *noise intensity*

spectrum simply the dimension $[pA\sqrt{Hz}]$. In section 7.1 a list of publications is summarized that demonstrates how different the words 'noise power spectrum' are interpreted by various authors. Sometimes it refers to the *single sided* noise intensity spectrum, while other authors meant *double sided* noise intensity spectrum. Both quantities are usually denoted with the symbol 'S', however they differ 3dB. This difference has been discussed in section 7.1.1. As a result, it is very difficult to inter-compare the noise performance of various reported lightwave receivers. In this text, *single sided* noise intensity spectra are specified.

Ratio noise measurements with synthetic noise

Instruments for reliable electrical noise measurements are commercially available, however this holds not for lightwave noise measurements. In 1991, we proposed a novel ratio noise measurement method for lightwave receivers [801,803], using lightwave synthetic noise. This lightwave source is discussed in chapter 9.

Figure 8.18 illustrates the proposed measurement setup.

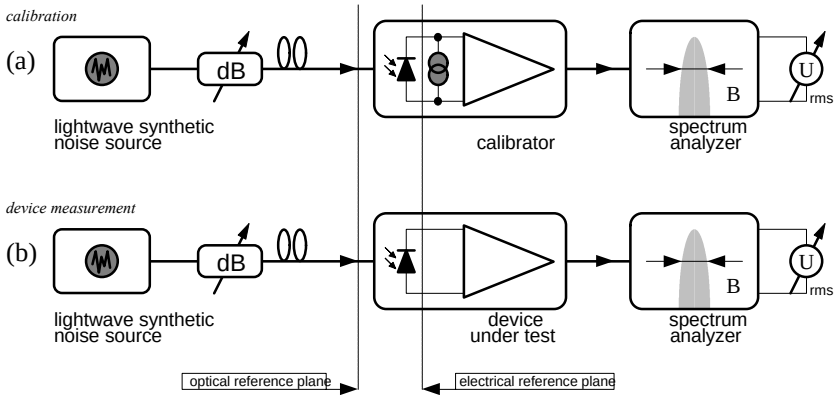


Fig 8.18 Block diagram of equivalent noise measurement on lightwave receivers. The synthetic noise is calibrated in (a) and then used in (b) to measure the ratio between the calibrated synthetic noise and the unknown device noise.

The first step in figure 8.18a is the calibration step, using one of the calibration procedures that are proposed in the section 8.3.2 and 8.3.3. In reference [801] shot-noise calibration with an incandescent lamp was used.

The calibration provides the spectral noise current ratio $\mu=(\sqrt{S_i})/(I_d)$, which is an inherent constant of the synthetic noise source. Subsequently, the optical power is adjusted to the desired power level for receiver noise measurement, using the lightwave attenuator. This optical power P_o is eventually measured with an optical power meter. When the photo-diode in the calibrator is also calibrated for optical power measurements, the optical power calibration can be integrated with calibration of the spectral noise current ratio μ .

The second step in figure 8.18b is the device measurement step, using calibrated synthetic noise to illuminate a receiver under test. Two reference planes are to be considered for noise specification:

- an optical reference plane at the optical input of the lightwave receiver, or
- an electrical reference plane at the output of the photo-diode.

The electrical reference plane is commonly used to specify the spectral *current noise density* $\sqrt{S_i}$ of the equivalent photo current (usually in pA/ $\sqrt{\text{Hz}}$). Its value is convenient in assessing the noise performance of the receiver pre-amplifier, but not of the lightwave receiver as a whole.

Noise specification observed from the optical reference plane is quite unorthodox, however it is more appropriate from a transmission system point of view. We will refer to it as lightwave *response noise density*¹¹ $\sqrt{S_o}$. Its level is proportional to $\sqrt{S_i}$ via the optical detector responsivity¹². The (scaled) dimension of $\sqrt{S_o}$ is [pW/ $\sqrt{\text{Hz}}$].

When P_{opt} is the optical power of the illumination, and when I_d is the dc photo current that is generated by that power, and when μ is the calibrated spectral noise current ratio of the synthetic noise source, then is the generated synthetic noise equal to:

$$\begin{aligned}\sqrt{S_o} &= \mu \cdot P_{\text{opt}} && \text{observed from the optical reference plane} \\ \sqrt{S_i} &= \mu \cdot I_d && \text{observed from the electrical reference plane}\end{aligned}$$

The optical power P_{opt} is always available for measurement, however the photo current I_d not. It is an internal signal of the device under test and requires receiver modification to make it external available. This is a significant advantage of specifying the lightwave response noise density. Furthermore, the synthetic noise is a white spectrum, observed from an *optical* reference plane. As a result, the unknown detector responsivity affects the spectral envelope of the synthetic current noise density $\sqrt{S_i}$. This illustrates that response noise density measurements are preferred to current noise density measurements.

Experiment

To demonstrate the reliability of ratio noise measurements with multi-level noise extraction, the equivalent input noise was measured of a lightwave receiver. Figure 8.19 shows the resulting spectra, observed from an output and an input reference plane. The output spectra in figure 8.19a represent the four reference output noise levels, as defined in section 8.4.1. They are the amplified levels of the combined device noise, synthetic noise and system noise.

The output spectra are frequency dependent since the receiver gain varies with the frequency. An additional ripple in the output spectra originates from mismatch errors between receiver and spectrum analyzer. These ripples change with the length of the linking cable, and will change when the receiver gain is measured with another instrument. This indicates that it would have been very difficult to correct for these mismatch effects when using direct noise measurements.

¹¹ The square S_o of the response noise density equals the noise equivalent power (NEP), normalized to 1 Hz noise equivalent bandwidth.

¹² The maximum responsivity of an arbitrary photo detector is restricted by physical constants: $r \leq (q\lambda)/(hc)$. For $\lambda=1550\text{nm}$, 1300nm and 850nm these limits are 1.250A/W, 1.049A/W and 0.686A/W respectively.

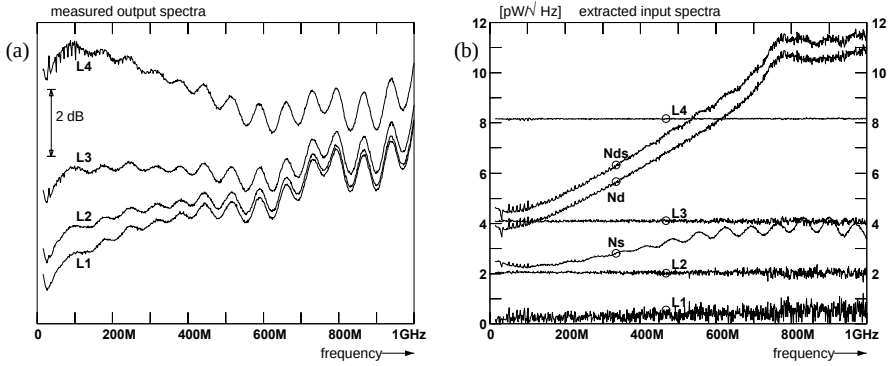


Fig 8.19 Measured noise spectra on electrical output (a) and optical input (b) of a lightwave receiver. The spectra L1 to L4 in plot (b) denote the levels of the synthetic noise source that are used as reference. The spectra L1 to L4 in plot (a) denote the detected output noise when the device is illuminated with these reference levels. The spectra N_d and N_s in plot (b) are the extracted device noise and system noise levels, observed from an optical reference plane. N_{ds} is the total equivalent input noise, resulting from the combination of N_d and N_s .

The equivalent input spectra in figure 8.19b represent various extracted noise levels, observed from an optical reference plane. Level L1 to L4 are the white synthetic noise levels, and were scaled from a calibrated level, using dc photo currents proportional to the illumination power. They were calculated from a multi-level extraction of virtual gain G and equivalent input noise $N_{ds} = \sqrt{S_{ds}}$, as defined in section 8.4.3.

All synthetic noise levels were equally weighted, assuming equally relative accuracy of the output spectra in figure 8.19a. As a result, the relative accuracy of the *extracted* synthetic noise levels degrades for very low illumination levels (note level L1 in figure 8.19b). These are realistic consequences.

The robustness of the multi-level extraction method has resulted in complete disappearance of the mismatch ripple in the spectra that originates at the input. This does not hold for the system noise $N_s = (\sqrt{S_s})/A$, since that noise is generated at the receiver output. Figure 8.19b demonstrates that the levels N_s and N_{ds} have some ripple, however it has disappeared in the device noise level $N_d = \sqrt{S_d}$. These are realistic results.

As a result, these experiments demonstrate the validation of the new multi-level extraction method, as proposed in section 8.4.3. The relative accuracy of the proposed noise measurement method on lightwave receivers is superior to the commonly used direct noise measurement. Conventional methods are usually unable to eliminate ripple effects that do not belong to the equivalent noise that is measured.

The absolute accuracy is mainly restricted by the accuracy of a common scaling factor. This means that all extracted levels are in equal proportion too high or too low. The uncertainty of dc optical power measurements takes the greater part of this (small) scaling error.

Furthermore, the proposed method provides gain information simultaneously, can be carried out simply, and results in noise measurements observed from an optical reference plane.

8.4.5. *Alternative lightwave receiver noise measurements*

Additional gain measurements facilitate the transformation of the equivalent input noise from the optical to the electrical reference plane and reverse. The property that our ratio measurement method extracts the noise at the *optical* reference plane, without additional gain measurements, is a non-trivial property. Most alternative ratio measurements extract the noise at an electrical reference plane. To demonstrate various superior performance aspects, our ratio measurement method is compared with alternative ratio methods.

(1) An alternative ratio method is electrical measurement of the four *noise parameters* of the receiver pre-amplifier. These parameters facilitate the reconstruction of equivalent input noise when the pre-amplifier is combined with the photo-diode. Although they require modification of the receiver under test to get access to internal reference planes, they are applicable. Automated measurement setups, using 50Ω noise sources and impedance tuners are commercially available.

⇒ We assume that these cumbersome and sophisticated methods suffer from poor accuracy performance. The output impedance of photo-diodes is relatively high, which enables the design of very low-noise pre-amplifiers. This causes that the equivalent input noise is often significantly lower than the thermal noise of 50Ω resistors. Since this level is significant lower than the cold noise level of 50Ω measurement methods ($18\text{pA}/\sqrt{\text{Hz}}$), we assume that the noise extraction will be inaccurate.

(2) A variant to the previous ratio method originates from using high-impedance *thermal noise* sources. Insertion and removal of various resistors, shunted to the photo-diode, may facilitate the extraction of equivalent input noise from thermal noise levels. The intention of the method is to avoid removal of the photo-diode and to avoid accurate impedance measurements.

⇒ At the beginning of our study to noise measurements, some experiments with this approach were performed, yielding poor results. The insertion of shunt impedances to the input affects the amplifier gain. Furthermore the equivalent device noise change with this impedance. This is a relevant problem, since the insertion of sufficiently high resistors is associated with noise levels that are too low for convenient noise measurements (see figure 8.20). Using four or more impedance values facilitates the extraction of all four noise parameters. This requires, however, the insertion of real as well as complex impedances. The RF-impedance of pure resistors is known from dc measurements with fair accuracy, however, insertion of very small capacitors requires impedance measurements at well-defined reference planes. This annihilates the simplism of the method, and requires an approach as is discussed in (1).

(3) The use of *shot-noise*, generated by an incandescent lamp, is well applicable under restricted conditions. This method is very similar to the shot-noise calibration previously

illustrated in figure 8.8. The calibration and the measurement, however, are simultaneously performed in the receiver under test.

⇒ This approach is restricted to receivers that gives access to internal photo currents for extracting the shot-noise level from dc measurements. Furthermore the method is restricted to receivers with open or connectorized photo-diodes. Pigtailed¹³ photo-diodes are hardly applicable, since that fiber is an impediment for adequate coupling of incandescent light.

Shot-noise methods require bright illuminations to generate adequate shot-noise levels. The associated dc photo current may overload the receiver under test by impeding the receiver biasing. Figure 8.20 demonstrates the superiority of synthetic noise, compared to shot-noise, on generation of dc offset currents. Shot-noise calibrators (see section 8.3.3) may be optimized for these high offset currents, since their bandwidth performance is irrelevant. This does not hold for all dedicated low-noise receivers under test, which restricts the application of the method.

	$\sqrt{S_i}$	=	1 pA/ $\sqrt{\text{Hz}}$	3 pA/ $\sqrt{\text{Hz}}$	10 pA/ $\sqrt{\text{Hz}}$
thermal noise	$R = 4kT/S_i$	=	16.4 k Ω	1.82 k Ω	164 Ω
shot-noise	$I_{dc} = S_i/(2q)$	=	3.12 μA	28.1 μA	312 μA
synthetic noise (B=50 GHz)	$I_{dc} \approx \sqrt{(2B \cdot S_i)}$	≈	0.32 μA	0.95 μA	3.16 μA

Fig 8.20 Comparison of the resistors or the dc currents that are associated with the generation of thermal, shot or synthetic noise. Synthetic noise combines a minimum in dc current with a maximum in noise level.

(4) Using *shot-noise* in combination with *balanced receivers* is also well applicable under restricted conditions. This method is very similar to the shot-noise calibration previously illustrated in figure 8.10. The balanced configuration avoids dc offset currents, since the photo current supplied by one diode is drained away by the other.

The method enables the use of LED and laser sources, to avoid coupling problems in pigtailed receivers. Care must be used in this measurement technique to obtain reliable results. The balancing must adequately suppress the undesired RIN below the shot-noise level to facilitate reliable shot-noise calibration. Kaspar et al. [805] reported the use of a laser source and Machida and Yamamoto [806] an LED source.

⇒ Although the method is very well applicable for balanced receivers, it requires access to internal photo currents for extracting the shot-noise level from dc measurements. The method is not applicable to other type of receivers.

(5) The use of *intensity noise*, that is modulated on the output power of a light source, is an alternative for synthetic noise. Explicit electrical modulation may be omitted when the RIN levels of lasers or LED's are adequate. This approach is very similar to our synthetic noise method, since it requires additional calibration and is applicable to a

¹³ For the ease of connecting to a fiber, some manufacturers supply LEDs, lasers and photodiodes with a short length of fiber, called pigtail.

wide range of receiver classes. It is the first method in this list that facilitates noise specification at an optical reference plane, without additional gain measurements.

⇒ The weak point of this approach is its calibration, since the spectral envelope of intensity noise is not white. The RIN curve of lasers is related to the electrical to optical transfer function and increases with the frequency up to the relaxation frequency of the laser. Typical values comprise several dB's up to a few GHz.

As a result, the calibration is required over the full frequency band of interest, which introduces many uncertainties due to unknown frequency responses in the calibration setup. Synthetic noise is white, up to tens, probably hundreds of GHz, which enables a simple calibration at relatively low frequencies.

(6) Our synthetic noise source is not the only lightwave signal, that generates white noise when it illuminates a photo-diode. Other examples are lightwave signals originating from *spontaneous emission*. Eichen et al. [816] proposed bandwidth measurements on lightwave receivers using amplified spontaneous emission from semiconductor optical amplifiers.

⇒ Whatever spontaneous emission source is used, their enormous spectral width makes them inferior to synthetic noise, with respect to dc-offset currents. A synthetic noise generator concentrates all available power in a bandwidth B that is selected by the user (see section 9.2.3). The more this power is spread out over a huge bandwidth, e.g. more than $5 \cdot 10^{12}$ Hz, the lower the usable spectral density will be. Measurements up to 100GHz would utilize only the first 2% of the width [816]. Since photo currents are proportional to the illuminated power, the associated dc photo current is significantly higher than was associated with synthetic noise for equal noise level.

The performance of spontaneous emission signals, with respect to dc offset current, is somewhere between the performance of synthetic noise and shot-noise.

8.4.6. Conclusions

In conclusion, a novel algorithm is developed for extracting equivalent input noise from multi-level noise measurements. This approach reduces random measurement errors by redundancy. Our algorithm minimizes the *relative* measurement error in a least-squares sense.

A new measurement method is proposed for measuring lightwave receiver noise. The method uses a new device, a synthetic noise source, a new multi-level extraction algorithm and dedicated calibration techniques to perform reliable ratio noise measurements over very wide frequency bands.

The high relative accuracy of our method is demonstrated. The absolute accuracy is mainly restricted by the accuracy of a common scaling factor. The uncertainty of dc optical power measurements takes the greater part in this (small) scaling error, followed by the accuracy of the calibrated noise level in the calibrator.

Our method is compared with various alternative methods for demonstrating its superior performance. The proposed method facilitates specification of noise at an *optical* input reference plane. It can be performed simply and is applicable to a wide range of lightwave receiver types.

8.5. Noise parameter measurements

A practical noise measurement setup requires the connection of a noise source to a DUT (device under test) via a transmission line, such as coaxial cables, wave guides or fibers. The previous section was restricted to sources that absorb all reflections to the DUT. As a result, the noise source was virtually located in any reference plane between source and DUT. A location shift of the reference plane requires no more than an amplitude correction of the source level due to (minor) transmission line losses. This simplified the noise extraction from the measurements significantly.

When the source admittance changes, the source becomes mis-matched. A location shift in reference plane will then require correction for source level as well as source admittance. Furthermore, the input excess-noise varies with the source admittance too. This makes that the measurement is preferably performed with a noise source that has a similar source admittance. As a result the noise extraction from the various measurements is complicated when the source admittances are mis-matched.

These difficulties are characteristic for electrical applications, and are not relevant for common optical applications. When a similar source admittance is not available, the preferred approach is different. All four noise parameters of the DUT have to be measured and the noise for the requested source admittance have to be reconstructed from these parameters.

This section starts with the state of the art in two-port noise parameter measurements. Various improvements will be summarized that result from replacing conventional $50\ \Omega$ noise sources with our noise-tees in state-of-the-art noise measurements.

A novel algorithm will be proposed that generalizes two-port noise parameter extraction of a device under test, when its input impedance is known. The algorithm holds for any combination of noise sources, each with arbitrary effective 'temperature' and impedance. When more than five sources are used, the algorithm extracts the most plausible solution, by minimizing the *relative* error.

This section demonstrates that our opto-electronic methods and algorithms are capable of noise measurements with distinct (mis-matched) source admittance, and of extraction of all two-port noise parameters.

8.5.1. State-of-the-art measurement methods

Most noise parameter measurements rely on noise figure measurements, on historical grounds. Its definition is related to the noise power delivered to the output, when a *device under test* (DUT) amplifies the thermal noise of a passive source admittance. By definition [817] it is the ratio between (1) the actual output power and (2) the output power when the DUT would have been noise free. When $F=1$, the DUT is noise free; when $F=2$, the input noise equals the thermal noise of the source admittance. As a result, the equivalent input noise spectra, that are associated with that resistance¹⁴, have the following intensities:

¹⁴ R is a chosen reference resistance, usually $50\ \Omega$, and T is a reference temperature. The IRE standard reference temperature [817] is defined as $T=290\text{K}$.

$S_{i,hot}$	$= (4kT/R) \cdot F$	= hot noise level of equivalent input noise current
$S_{u,hot}$	$= (4kT \cdot R) \cdot F$	= hot noise level of equivalent input noise voltage
$S_{i,excess}$	$= (4kT/R) \cdot (F-1)$	= noise current, in <i>excess</i> to the thermal noise of the source
$S_{u,excess}$	$= (4kT \cdot R) \cdot (F-1)$	= noise voltage, in <i>excess</i> to the thermal noise of the source
$S_{i,cold}$	$= (4kT/R)$	= cold noise level due to thermal noise current of the source
$S_{u,cold}$	$= (4kT \cdot R)$	= cold noise level due to thermal noise voltage of the source

The noise factor (or noise figure), is convenient in a 'transmission line environment' where all devices are designed for operation at equal source and load admittances. The definition is meaningless when the admittance (or impedance) of the input termination is not specified. To our opinion, the use of noise factor F for 50Ω is very inconvenient when amplifying signals generated by ideal voltage or current sources (such as photo-diodes).

The measurement of noise factor F at a distinct source admittance requires a calibrated noise source connected to the DUT. A tunable network between source and DUT facilitates the adjustment of the admittance of interest. Then the output noise is detected, and F is measured via ratio methods. This method is often referred as the Y-factor method, probably due to formula 10 in [821]. The use of automated noise figure meters, that controls the 'hot' and 'cold' states of the noise source, is more preferred.

The noise factor F varies with source admittance Y , and Rothe and Dahlke [713] published in 1956 a simple four-parameter expression¹⁵. For a nice overview of this theory, see Hartmann [722] with many relevant references.

At a specific source admittance (Y_{opt}), the noise factor is minimal (F_{min}). Most noise parameter measurements are focused on the extraction of these parameters, and an additional scaling factor R_n : [F_{min} , $Re(Y_{opt})$, $Im(Y_{opt})$, R_n] (see chapter 7). These four (real) parameters can be transformed into any form of interest, using the theory of chapter 7.

Paired hot and cold noise measurements, with search to F_{min}

- (1) • The classic approach, recommended since 1960 by the IRE [818,819,820,821], uses a loss free *tuner* (a variable-ratio transformer or a movable stub tuner). At each admittance of interest, the noise figure is measured with hot/cold ratio measurements. The admittance is adjusted, using trial and error methods, to find a minimum readout F_{min} for the noise figure. The associated source admittance provides a specific value Y_{opt} . The value R_n is then calculated from a previously measured pair of (F , Y), or extracted from all measured pairs (F , Y) using a graphic linear regression method.
- Lange [822] proposed in 1967 an alternative noise parameter $N \stackrel{\text{def}}{=} R_n \cdot Re(Y_{opt})$, as replacement for R_n . He demonstrated that N as well as F_{min} are invariant under lossless transformation. Y_{opt} transforms through a lossless two-port like any admittance. This illustrates that F_{min} and N are fundamental properties of the DUT in transmission line operation. Therefore the admittance Y_s can be measured at the

¹⁵ Noise figure: $F = F(Y) = F_{min} + \frac{R_n}{Re(Y)} \cdot |Y - Y_{opt}|^2 = F_{min} + N \cdot \frac{|Y - Y_{opt}|^2}{Re(Y) \cdot Re(Y_{opt})}$

most convenient reference plane located between the two-port under test and the noise source, with no effects on N and $F_{\min}$. They are terminal invariant in a lossless region, and Y_s is the only parameter to be transformed.

Paired hot and cold noise measurements, with curve fitting

(2) The drawback of method 1 is that the tuner must provide all admittance values of interest. The optimal admittance was found on a trial and error base, guided by manual curve fitting techniques.

- Lane [823] was the first to introduce in 1969 an algebraic fitting approach on measured pairs of (F, Y) for discrete admittance values (four or more). He transformed that expression into a linear form to extract the desired parameters from an over-determined set of simultaneous equations. Caruso and Sannino [827] improved the accuracy in 1978, using properly chosen source admittance values. Both methods [823,827] minimize the absolute error ΔF in F in a least-squares sense, to improve the overall measurement accuracy (one-dimensional minimization).

- Mitama and Katoh [828] improved in 1979 the curve fitting method, using a three-dimensional minimization. They minimized the combined absolute error $(\Delta F^2 + \Delta Y^2)$ in an iterative approach.

Arbitrary hot and cold noise measurements, using input admittance.

(3) The drawback of methods 1 and 2 is that at least eight output noise measurements are required: four pairs of 'hot' and 'cold' measurements. Furthermore, method 1 and 2 require that the source admittance does not change with the effective 'temperature' of the source. More economical methods make use of additional information on the input admittance. This relaxes the requirement to perform the measurements in pairs with equal source admittance, and reduces the number of noise measurements.

- Engen [824] proposed in 1970 a method with four noise measurements and an input admittance measurement. He measured the output noise levels (P_{oh}, P_{oc}) associated with a matched hot/cold source, the output noise levels (P_{om}, P_{om}) associated with a sliding short adjusted at maximum/minimum output level.

- Adamian and Uhler [825,830] proposed in 1973 a method with five noise measurements and an input admittance measurement: one pair of hot/cold measurements with a noise source, and three additional cold measurements at arbitrary source admittance values. They considered that different admittances are allowed for the noise source in the 'hot' and 'cold' state, and suggested an iterative approach to use this information in the noise parameter extraction.

Davidson et al. [831] proposed in 1989 a non-iterative correction method for variation in source admittance when varying the noise 'temperature'. They demonstrated the advantages when the measurements are corrected for these effects. As a result, it is not necessary to measure cold noise power at the source admittance produced by the hot source.

Equal temperature noise measurements

(4) Probably the most convenient measurement method is a method that uses 'cold' measurement only, or have at least all sources at equal 'temperature'.

- Gupta [829] demonstrated in 1983 that such a solution does not exist when the output noise power is the only measured quantity.

Multi-port noise measurements, using N -port parameters

- (5) All previous methods use switched or tunable input impedances to extract the four noise parameters of the DUT. This is a disadvantage when the measurement frequency ranges over several decades .
- Wedge and Rutledge [833] proposed in 1992 a two-port method that performs simultaneous measurements of the noise waves that emanate from input and output ports of the DUT. The proposed setup uses two noise sources to perform ratio measurements on the noise waves originating from port 1, originating from port 2, and a mix of both waves using 90° and 180° hybrids. Transformation of all noise contributions to equivalent input values requires the measurement of the all two-port parameters of the DUT.

The multi-port noise measurement requires additional two-port parameter measurements to extract the *input* noise parameters. The distinct hot and cold noise measurement requires an additional input admittance measurement on the DUT. When the input admittance is *not* available for measurement, then the paired hot and cold noise measurements apply.

Automated methods, using hot and cold measurements, are commercially available. They use electrical noise sources (50Ω), in combination with automated tuners or switched admittances. Two-port tuners are commonly used for the paired hot and cold noise measurements, however additional circulators facilitate the application of one-port tuners.

8.5.2. Proposed improvements to state-of-the-art noise measurements

Most state-of-the-art methods of the previous section are reliable methods. Nevertheless, there is still room for improvement. This subsection summarizes various improvements that result from replacing conventional 50Ω noise sources with our noise-tees, and that result from our new noise extraction algorithms.

All reflective source

(-) The paired hot/cold measurements require noise generators with variable source admittance. Conventional methods (50Ω source, two-port tuner) must be content with limited tuning ranges for admittance values since high reflective sources (shorts, opens, offset shorts, etc.) are associated with very low noise levels.

(+) Our noise-tee is a *two-port* device, and requires no more than a one-port admittance tuner. An offset short on one port will result in a 100% reflective source for all frequencies, but will not block all available noise power. This makes our noise-tee wider applicable than 50Ω noise sources.

Invariant source admittance

(-) Paired hot/cold measurements require noise generators with source admittances invariant under effective 'temperature' variations. Section 8.2.4 demonstrated that commercially available sources are liable up to 10% variation of output admittance.

(+) The admittance of our noise-tee is highly insensitive to variation of noise level.

Zero cold level

(-) Paired hot/cold measurements, using conventional 50Ω noise sources, are associated with a minimum 'cold' noise level. The thermal noise of the cold 50Ω resistor restricts the measurement accuracy on very weak noise currents, especially when their level is significant lower than the 'cold' level of $18\text{ pA}/\sqrt{\text{Hz}}$.

(+) Our noise-tee enables measurements with loss-free source admittances. As a result, the minimum 'cold' level is zero.

Source removal not required

(-) Distinct hot and cold measurements perform 'cold' measurements with zero reference levels, by replacing the noise source with a loss-free admittance.

(+) Since our noise-tee is equivalent with a pure transmission line when not in use, there is not necessary to remove it for 'cold' noise measurements. This simplifies the measurement because it saves an additional action.

Generalized extraction algorithm

(-) The distinct hot and cold measurements found in the literature use one hot/cold measurement. They reported a suitable algorithm when the hot/cold measurement has equal source admittance [825,830], and an improved algorithm that corrects for parasitic variation errors when the admittance varies with the effective temperature [831]. This is an inconvenient approach when the method is generalized for several hot and several cold sources, all with different admittances.

(+) The succeeding subsections propose a novel algorithm for extracting noise parameters. They are generalized to facilitate noise extraction for sources, each with an arbitrary effective 'temperature' and admittance.

Minimization of relative errors

(-) State-of-the-art noise extraction algorithms use an over-determined set of equations, and are focused on extracting a plausible solution with minimum error. The algorithms that we found in the literature [823,827,828,832], focus on minimization of *absolute* errors of one or more *noise parameters* (usually the absolute error ΔF).

On the other hand, most applications use these noise parameters to predict the equivalent input noise for a specific source admittance. They require an optimal prediction of *that* value. The accuracy is preferably invariant with the source admittance value, or is optimal for the source admittance of interest. The optimization goal varies with the application. Device modeling applications require usually optimization of the *excess-level* of the noise current (or voltage), while amplifier applications prefer the optimization of the *hot-level* of the noise current (or voltage).

(+) The algorithms proposed in the succeeding subsection minimize *relative* errors. They minimize the equivalent input excess-noise current. Nevertheless, the train of thought is applicable to a wide range of goal quantities of interest.

The succeeding subsections will demonstrate that our opto-electronic measurement methods, with multi level noise measurements, are capable of full noise-parameter extraction.

8.5.3. Input noise measurements for mismatched source admittances

When designing low noise amplifiers for specific sources then the equivalent noise associated with the specified source admittance is of particular interest. Measurement of this noise level facilitates the selection of optimal devices and their biasing, having noise-optimal performance for that specific source.

These type of measurements require device noise measurements, in which the source admittance is simulated with an admittance tuner. The two-port noise-tee configuration of section 8.2.4 enables the reproduction of these circumstances with an additional one-port admittance tuner. Figure 8.21 illustrates the basic measurement setup and the various mathematical transformation steps.

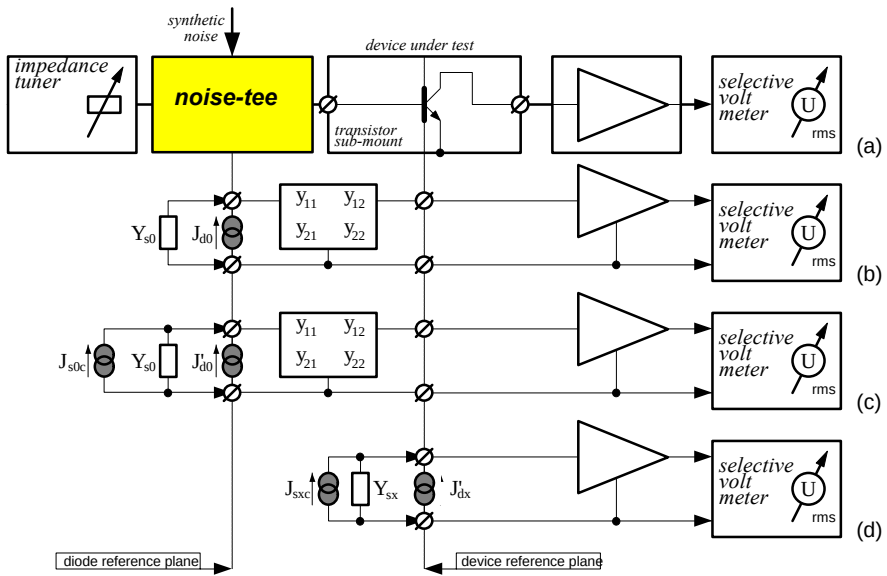


Fig 8.21 Basic measurement setup (a) for the measurement of device noise J_{d0} observed from the diode reference plane for specified admittance Y_{s0} . In (b,c,d) the various steps are illustrated to transform this noise (including thermal noise effects) to the device excess-noise J'_{dx} observed from the device reference plane (excluding thermal noise effects J_{sxc}).

Taking it roughly, the measurement and the extraction algorithm consist of the following succeeding steps:

- (1) Measurement of the *hot noise level* J_{d0} of the device noise current, observed from the diode reference plane of the noise-tee, for specified admittance value. An admittance tuner is used to realize this source admittance.
- (2) Extraction of the *excess noise level* J'_{d0} , of the device noise current, by subtracting the *cold noise level* (J_{s0c}) due to thermal noise effects in the tuner admittance.
- (3) Transformation of this excess-noise from the diode reference plane to the device reference plane: J'_{dx}

- (4) Addition of thermal noise J_{SXC} when the *hot noise level* of the input noise current is requested, in stead of the excess noise level.

The first step is similar to the ratio measurement on lightwave receivers, as described in section 8.4.4. The other steps are rather similar to the transformation of calibrated noise levels to arbitrary reference planes, as described in the sections 8.3.4 and 8.3.5. The transformation steps and the extraction algorithm for current J'_{dx} are described in detail in appendix L.

To demonstrate the feasibility of our extraction method, the device excess-noise of a bipolar junction transistor are measured for various admittance values. We realized a simple switched admittance tuner with resistors and capacitors, and measured the device excess-noise for six arbitrary settings of the tuner. For a detailed description of this experiment, see Mahmoudi [834].

The device noise was extracted from ratio measurements using four different reference levels. The most plausible device noise was extracted from these values, and then transformed to the device reference plane, using measured two-port parameters and source admittance values. Figure 8.22b shows the extracted device excess-noise current, measured for the source admittance values of figure 8.22a.

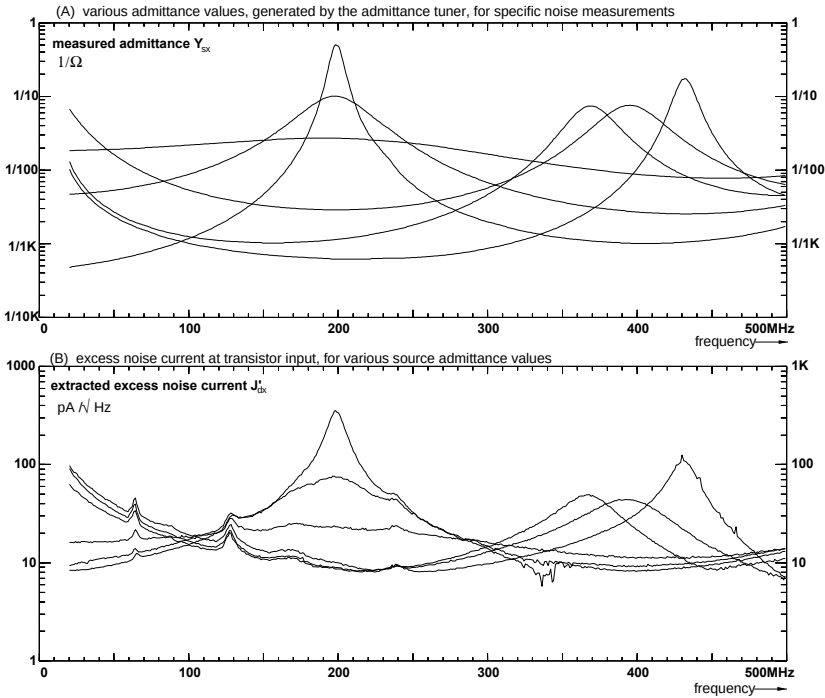


Fig 8.22 Various measured source admittance (a) values Y_{SX} and their associated amplitude spectra (b) of the device excess-noise current J'_{dx} . These measurement are observed from the device reference plane, as illustrated in figure 8.21.

The experimental admittance tuner was composed of resistor-capacitor pairs, shunted to a common 50Ω microstrip transmission line. Manual on and off switching was used to

vary the overall admittance. The tuner, the noise-tee as well as the transistor fixture were constructed from standard epoxy printed circuit boards. The average distance between tuner and device reference plane was approximately 14 cm.

This relative long transmission distance between tuner and DUT has caused the resonant behavior of admittance Y_{sx} . As a result, arbitrary admittance values are restricted to small frequency bands, especially for relatively high and low admittance values (compared to 50Ω). Furthermore this distance has restricted the dynamic range of tuner admittance values for frequencies around 100MHz and 300MHz, observed from the device reference plane. Proper construction will improve the tunability of the source admittance over a wider frequency band.

8.5.4. Extraction of device noise-parameters

This discussion in section 8.5.3 illustrates one can't position the tuner adjustment of a practical setup to *any* admittance of interest. Therefore another approach is required. It requires the prediction of the noise at specified source admittance, based on noise parameter measurements using four or more mis-matched sources. Figure 8.23 illustrates this supplementary approach.

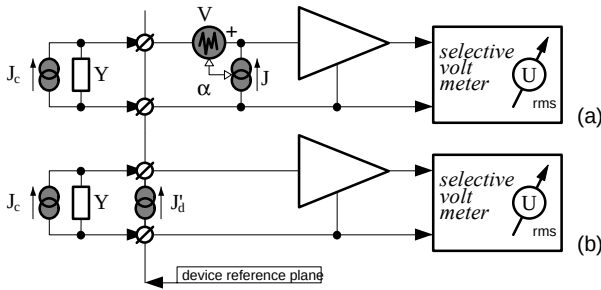


Fig 8.23 The device excess-noise current J'_d for arbitrary source admittance values Y is obtained from a full noise representation of the input port of the device under test. Current J_c is the thermal noise of the source admittance Y ('cold' noise).

Figure 8.23a shows a full noise representation of the device under test, using a correlated pair of a noise voltage source and a noise current source. This model is applicable to any source admittance of interest. Figure 8.23b shows an equivalent representation. Its validation is restricted to the chosen value of the source admittance. The device excess-noise current J'_d is calculated from Y , V , J and their correlation α . Let S_J and S_V be the (self)-intensity spectra of $J(t)$ and $V(t)$ respectively, let S_{JV} be their cross-intensity spectrum, and let S_i be the (self) intensity spectrum of the equivalent excess-noise current $J'_d(t)$. As discussed in section 7.3, the following relation holds:

$$S_i = S_j + |Y|^2 \cdot S_v + 2 \cdot \text{real}(Y) \cdot \text{real}(S_{j|V}) + 2 \cdot \text{imag}(Y) \cdot \text{imag}(S_{j|V})$$

$$\text{let: } \begin{array}{ll} x_1 \stackrel{\text{def}}{=} S_j & x_3 \stackrel{\text{def}}{=} 2 \cdot \text{re}(S_{j|V}) = S_{j|V} + S_{j|V}^* \\ x_2 \stackrel{\text{def}}{=} S_v & x_4 \stackrel{\text{def}}{=} 2 \cdot \text{im}(S_{j|V}) = S_{j|V} - S_{j|V}^* \end{array}$$

$$\boxed{S_i = x_1 + |Y|^2 \cdot x_2 + \text{re}(Y) \cdot x_3 + \text{im}(Y) \cdot x_4}$$

This is a linear relation, with $[x_1, x_2, x_3, x_4]$ as noise parameters. It illustrates that the overall measurement goal has been focused on the extraction of $[x_1, x_2, x_3, x_4]$, to evaluate the excess-noise intensity S_i .

Note that our primary interest is the prediction of J'_d , and not the evaluation of the noise parameters. Therefore, we will restrict our self to a convenient (non-standard) representation format $[x_1, x_2, x_3, x_4]$ to evaluate the device excess-noise for arbitrary source admittances Y . Furthermore, optimal accuracy is required for the excess-noise and not for the noise parameters. Therefore, it is useless to find some optimal solution for $[x_1, x_2, x_3, x_4]$. Instead of that, the extraction algorithm must optimize the relative accuracy of S_i .

This work has resulted in two robust extraction algorithms to extract noise parameters from noise measurements. Appendices *M* and *N* describe both algorithms in detail.

- The first algorithm, in appendix *M*, requires information on the input admittance of the device under test. This algorithm is applicable with arbitrary hot and cold measurements. Variation of source impedance is permitted when its noise output level varies. This principle is known from the literature, however the algorithm has been generalized in this study. The algorithm applies for any set of detected noise levels and associated (arbitrary) admittance values.
- The second algorithm, in appendix *N* does not require any information on the device input impedance. The consequence is that hot and cold measurements are to be performed in pairs with equal source admittance. Another consequence is that more independent levels are to be detected: at least eight (2×4) noise levels instead of five with the previous method. This principle is also known from the literature, however this study has supplemented it with multi level measurements and applied it to sources with noise-tees.

We applied [834] the second algorithm to the extraction of $[x_1, x_2, x_3, x_4]$ from the measurements of figure 8.22. We observed that these noise parameters are applicable for predicting input noise levels for specific source impedances. This is demonstrated in section 8.5.3 for resistive, capacitive and inductive sources.

Additionally, we transformed $[x_1, x_2, x_3, x_4]$ into autonomous noise parameters, as proposed in section 7.3.3, for extracting the parameters of the device noise model discussed in section 7.3.3. We observed that these transformed parameters are quite sensitive to small measurement errors in $[x_1, x_2, x_3, x_4]$. These preliminary results indicate the high demands on measurement accuracy and reproducibility when extracting two-port device models from noise parameter measurements.

8.5.5. Extraction of input noise level for arbitrary source admittances

To demonstrate the feasibility of our second extraction method (see section 8.5.4 and appendix *N*) with paired (multi-level) hot/cold measurements, the measured [834] results

from figure 8.22 are used to extract the noise parameters of that device. These four parameters are extracted from six independent measurements, each with four reference levels.

From these parameters $[x_1, x_2, x_3, x_4]$, the equivalent excess-noise was predicted for pure capacitive sources, pure inductive sources and pure resistive sources. Figure 8.24 shows the predicted values of the equivalent excess-noise intensity spectrum S_i .

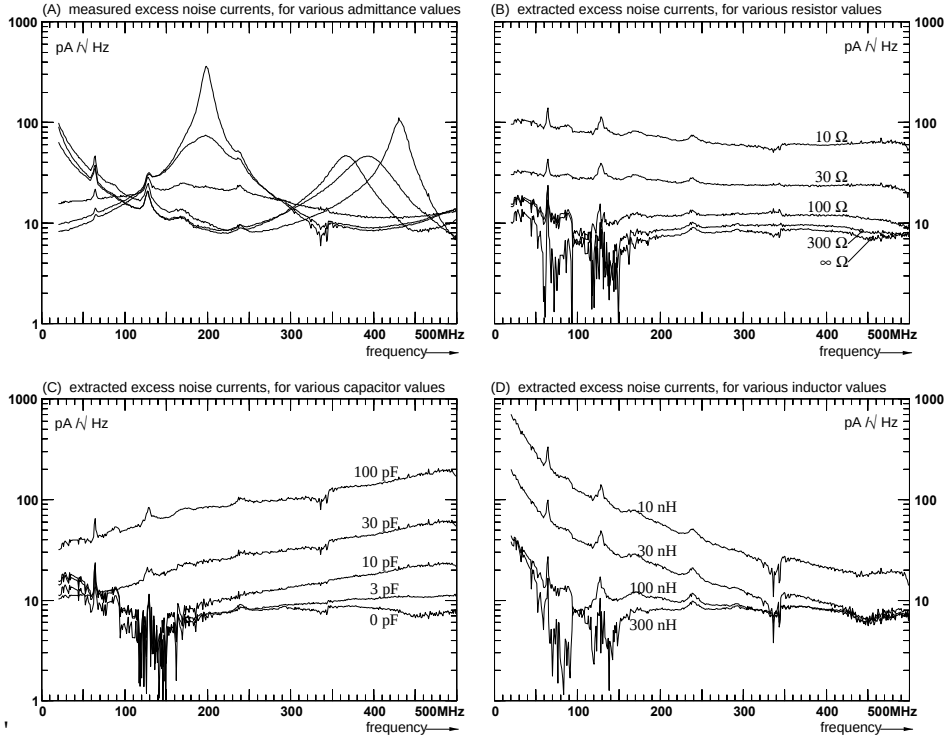


Fig 8.24 Extracted excess-noise current, of a bipolar junction transistor (BFR92a, $U_{ce}=10V$, $I_c=10mA$, $I_b=0.25mA$), observed at its input. The predictions are based on the extracted noise parameters, using the measurements of figure 8.22.

The excess-noise is the equivalent input noise, minus the thermal noise of the source admittance. It changes with the admittance value.

Figures 8.24b, 8.24c and 8.24d show the reconstructed excess-noise levels S_i , reconstructed for ideal resistive sources, ideal capacitive sources and for ideal inductive sources. They illustrate that the equivalent excess current noise reduces with decreasing admittance. This noise is minimal for infinite admittance values. A significant decrease is not observed above $|Z|=300\Omega$. These plots are very useful to predict the equivalent input noise for sources with these source impedance values.

Figure 8.24a shows similar plots, associated with admittance values Y that were used for the extraction process (shown in figure 8.22a). As expected, their values match very well with figure 8.22b, which illustrates the validation of our extraction method.

The above examples evaluate the noise parameters as intermediate result. Nevertheless, these parameters can be transformed into any format, e.g. as autonomous noise parameters as discussed in section 7.3.3. This facilitates the development and assessment of adequate transistor noise models.

These noise parameters are extracted, as reported by Mahmoudi [834]. The accuracy, however, of these experimental results is inadequate to perform reliable model extraction. The spikes and dips in the extracted values, shown in figure 8.24, are indicative for the high accuracy and reproducibility that is required for these types of measurements. As a result, improvements of the measurement are aspects for further investigations.

8.5.6. Conclusions

In conclusion, various state-of-the-art methods are summarized to measure the input noise parameters of a device under test. The difference between the methods is characterized by the need for two-port measurements, the need for one-port measurement or no transfer measurement at all.

Several improvements are discussed that may result when applying our noise-tee to state-of-the-art methods.

It has been demonstrated that our opto-electronic measurement methods, with multi level noise measurements, are capable of full noise parameter extraction. Novel algorithms are developed for extracting noise parameters. They are generalized to facilitate noise extraction for sources, each with an arbitrary effective 'temperature' and admittance.

Chapter 9

Lightwave synthetic noise generation

The characterization of the noise performance of lightwave receivers is of basic importance for optical communication systems. Spectral intensity measurements on noise are conveniently be made with signal generators delivering white noise with known noise levels. These methods are superior in simplicity and accuracy, compared to direct noise measurements (see section 8.1), since additional gain measurements are not required. Other methods, including direct methods, require a separated measurement of receiver gain to reconstruct the input noise from the measured output noise.

As early as 1953, the IRE recommended [817] the application of white noise generators (dispersed signal sources) in electrical noise measurements. A convenient and commercially available *lightwave* noise generator is lacking (1994), and therefore noise measurements on lightwave receivers are usually restricted to direct measurements. Recently, we proposed a lightwave synthetic noise generator for noise measurements that generates white noise currents when illuminating PIN photo diodes [802,801,803,901,902]. The proposed generator is primary intended for electrical and optical measurements of noise, however it is also applicable for transfer measurements. In this chapter, we analyze the principles of synthetic noise generation in detail. Much effort of this thesis work was focused on the development of such a noise generator.

State of the art of lightwave based noise measurements

In the literature, various lightwave methods are described to measure noise in lightwave receivers. These methods are listed below.

- Direct noise measurements on lightwave receivers are widely used [807, 808, 809, 810, 811, 812, 813, 814, 815]. In these references receiver output noise was measured directly (usually with a spectrum analyzer), then receiver gain was measured or estimated, and the equivalent input noise was reconstructed from these quantities. These methods suffer from limited accuracy since the separated gain measurement must accurately incorporate all mismatch errors and all noise detection errors of the noise measurement.
- Shot noise methods are more convenient, since the shot noise level is known from dc measurements: $S_i = 2q \cdot I_{dc}$. Kaspar et al. [805] compared the equivalent input noise of a *balanced* receiver with the white shot noise generated by a laser. The balanced configuration facilitates suppression of (unknown) laser RIN. Machida and Yamamoto [806] measured squeezing of light using a similar ratio noise measurement. They generated white shot noise in a balanced pair of photo diodes using an LED. Clean light sources, such as incandescent lamps [801], are sometimes preferred since it is assumed that they do not suffer from unknown RIN levels. All shot noise methods suffer from large dc photo currents to generate an adequate shot noise level. This may overload unbalanced receivers.

- Amplified spontaneous emission from optical amplifiers [816] may provide higher noise levels for comparative dc currents. Nevertheless, the noise power is spread out over thousands of GHz, which indicates that most noise power is unused.
- Lasers modulated with white electrical noise or lasers with high RIN provide higher noise levels. They are not white over a wide frequency band, which prevents them from a simple wideband calibration, performed at low frequencies only.
- We proposed *synthetic noise* for lightwave receiver noise measurements [801,803] and for electrical noise measurements [802]. We used the source of Wang et. al. [907], who proposed a bandwidth measurement setup, improved its spectral flatness [901,902] and calibrated it with shot noise [801]. The major advantages of synthetic noise are that:

(1) the total rms-noise current is as high as 70% of the dc photo current, and that

(2) all available noise power is concentrated in a user definable bandwidth B.

A variable noise bandwidth is an advantage to maximize noise power and to minimize spurious response. The wider the spectral width of the signal is, the more the available power is spread out over the frequency band and the lower the spectral intensity will be. A large portion of the synthetic noise band is applicable for white noise purposes, and can be as wide as twice the maximum lightwave FM sweep of the laser. Variable noise band widths up to hundreds of GHz are assumed to be feasible when multi-section lasers are applied.

All these aspects make the synthetic noise method superior to the previous methods.

State of the art of synthetic noise generation.

As early as 1953, the IRE recommended [817] the use of white noise generators (dispersed signal sources) for noise measurements. Suggested white noise sources are (1) *natural* noise generators such as noise diodes or gas tubes and (2) *synthetic* noise generators such as oscillators whose frequencies are swept through the band at a uniform rate.

Electrical synthetic noise generators are inconvenient for wide band noise measurements, since their spectral width is restricted to the maximum FM sweep of the modulated oscillator. On the other hand, the lightwave synthetic noise principle is quite applicable for the generation of wideband noisy signals. This is because a weak FM modulation at lightwave frequencies will easily cause a frequency swing of several GHz. Wang et al. [907] proposed in 1989 a lightwave source for bandwidth measurements, using the delayed self-homodyne spectrum of a DFB laser. The delayed self-homodyne principle [903,905], and more recently the self-heterodyne [904] principle, are well-known from laser linewidth measurements [903,905]. In addition, Wang et al. modulated the laser frequency with a sinusoidal signal to spread out the homodyne spectrum over a very wide frequency band. They recommended a delay significantly larger than the coherence length of the DFB laser to smooth the output spectrum into a white noise spectrum. As a result, the Wang experiment [907] required a delay of more than 1 km fiber length.

A first prototype of this new lightwave setup, that was made available for this thesis work, was developed by van Deventer [908]. We developed novel techniques for wideband noise measurements on lightwave receivers [801,803 and section 8.4.4 and 8.3.3], using this lightwave source, and proposed the name *synthetic noise source*. We proposed this principle for a novel electrical noise source [802], and demonstrated some advantages, compared to conventional electrical sources. In addition we proposed the noise-tee [802] as discussed in section 8.2.

Next, we improved the spectral flatness of synthetic noise using electrical noise injection and triangular modulation [901,902] and we proposed multi-loop stabilization. The noise injection facilitates the use of multi section lasers. They are preferred rather than DFB lasers due to their superior FM modulation performance. Since these lasers are commonly optimized for maximal coherence, the application of noise injection makes a delay length of several kilometers superfluous. Some improvements are listed in figure 9.1.

<i>Wang experiment</i>	<i>our improvements</i>	<i>remarks</i>
sinusoidal FM modulation	triangular FM modulation	This improves the usable portion of the available noise bandwidth
parasitic laser noise (laser linewidth)	external noise injection	This reduces the spectral ripple. Pink noise injection is preferred to white noise injection to make efficient use of available noise power
long delay required for adequate incoherence. (1km, combined with 100kHz modulation)	short delay is adequate (e.g. 10m, combined with 10MHz modulation)	Due to noise injection. Strong incoherence improves the spectral ripple, and noise relaxes the incoherence requirements such as delay length. A short delay length simplifies practical implementation.
DFB laser. Linewidth of 20 MHz or more is essential for good performance	multi section lasers are also feasible, linewidth is irrelevant	Due to short delay. Multi section lasers are superior in FM performance. This facilitates huge noise bandwidths
unstabilized	multi-loop methods for stabilization of noise intensity spectrum	Independent stabilization of optical power and noise intensity spectrum

Fig 9.1 Overview of various improvements that made a simple reproduction of the synthetic noise source viable.

Highlights of this chapter

The highlights of this chapter, which are developed during the work on this thesis, are:

- Development of a novel lightwave device, a *synthetic noise generator* that provides white noise when it illuminates PIN photo diodes.
- *Improvements* to lightwave synthetic noise generation, using noise injection and triangular modulation [901: patent pending, 902].

- Dual loop method, for independent *stabilization* of the synthetic noise level as well as the optical power of synthetic noise generators [901: patent pending].
- Detailed *analysis* of many practical and theoretical aspects of synthetic noise generation.
- Practical implementation of a synthetic noise generator to demonstrate the feasibility of the proposed methods.

9.1. Basic principles of synthetic noise generation

A lightwave synthetic noise generator is an FM modulated laser, of which the lightwave spectrum is down converted into the 'electrical' frequency band. It uses the delayed self-homodyne principle for down conversion.

This introductory section describes the basic ideas behind synthetic noise generation, and demonstrates the feasibility and performance of this new noise source. The backgrounds are analyzed in section 9.2 from a practical point of view, and in section 9.3 from a theoretical point of view.

9.1.1. Delayed self-homodyne principle with FM modulation

The lightwave synthetic noise source starts from a laser source, of which the lightwave output is modulated in frequency. The block diagram in figure 9.2 presents this by means of an electrical source that generates a periodic signal with frequency f_m , such as a 10 MHz sinusoidal or triangular signal. The modulation is deep enough to modulate the lightwave frequency over several giga-Hertz.

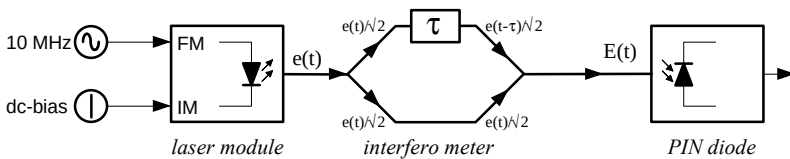


Fig 9.2 Block diagram of the synthetic noise generator. The modulation frequency modulates the lightwave frequency. The delay in the interferometer causes two lightwave FM-signals, that differ in frequency. The PIN-diode performs the duties of a down converting mixer, to generate an electrical differential frequency that sweeps periodically at modulation rate.

The use of a noise-free laser would have led to a modulated lightwave spectrum with discrete equidistant frequencies (see figure 9.3). The difference in frequency between each frequency component and the lightwave carrier is exactly a multiple of the modulation frequency f_m .

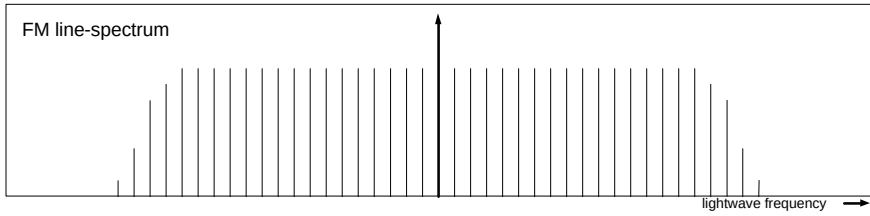


Fig 9.3 The laser output (FM) spectrum is essentially a line spectrum, with equidistant lines. This is a simplified picture; in practice the spectrum is generally not flat as is shown in figure 9.10.

The linewidth of practical lasers prevents the lightwave spectrum from being a perfect line spectrum. As a result, a lightwave spectrum analyzer with very high resolution will indicate a comb spectrum that is similar to the spectrum in figure 9.5a.

Next, the lightwave signal is split, one beam is delayed and both beams are combined by means of a fiber optic interferometer. The difference in length between the two optical branches ranges from a few meters to many kilometers, in practical setups.

The composite lightwave signal is essentially the composition of two FM modulated lightwave signals. Under optimal delay conditions, which are discussed in a little while, the momentaneous frequencies oscillate symmetrically around a common mean value (see figure 9.4). The frequency sweep depends on the FM modulation depth.

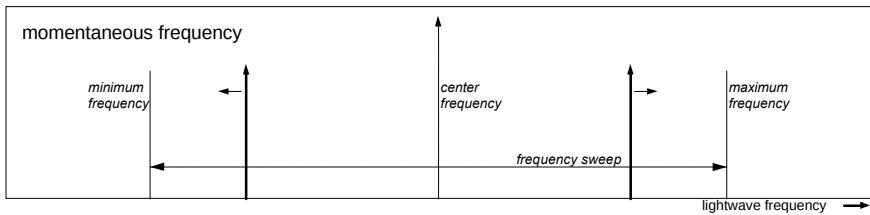


Fig 9.4 The output signal of the synthetic noise generator is essentially the composition of two FM modulated lightwave signals, of which the momentaneous frequencies oscillate symmetrically around a common mean value.

The composed lightwave signal illuminates the PIN photo diode. This diode performs the duties of a down converting mixer, because of its linear *power* response. Since this is similar to a *quadratic* field strength response, the non-linear transfer converts the two optic frequencies into components with other frequencies. The sum and differential frequency are included.

DC-currents and currents with the differential frequency are the principal currents that escape from the electrical passband of the diode¹. All other converted components have frequencies in the lightwave range and are therefore unable to escape from the PIN diode chip as electrical current.

¹ Note that spurious AM and shotnoise also cause minor contributions to the diode current.

The momentaneous frequency of the generated photo current equals the difference in lightwave frequency of the two lightwave signals. This difference changes periodically at the rate of the modulation frequency f_m , which causes the momentaneous frequency to change at the same rate. The maximum momentaneous frequency of the photo current equals the maximum frequency difference, which can be as high as twice the modulation depth. As a result, a weak modulation in frequency of the lightwave signal causes a giant absolute change in frequency of the photo current.

The concept of momentaneous frequency is adequate for a time-domain description, however, frequency domain descriptions require Fourier series. Since the modulation frequency of practical synthetic noise generators is 100 kHz or more, a spectrum analyzer is unable to track these rapid frequency changes. As a result, they will display the Fourier components of the photo current, with many equidistant frequencies. The detected spectrum is similar to the spectrum in figure 9.3, with the center frequency at zero Hz.

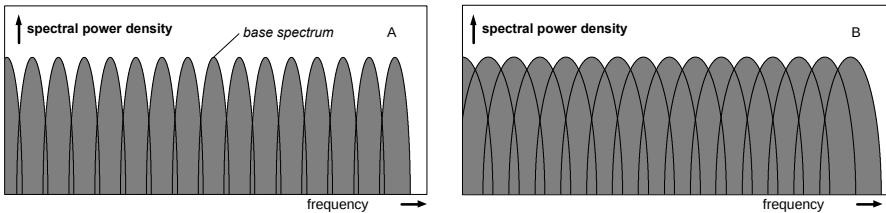


Fig 9.5 The output (FM) spectrum of the synthetic noise generator is essentially a line spectrum, with equidistant lines. The phase noise (linewidth) of each line fades the discrete spectrum into a continuous spectrum. The envelope ripple of spectrum (a) is deeper than that of spectrum (b) because the width of the individual lines in (b) is wider. The wider the linewidth is, the more a white noise spectrum is imitated.

The non-zero linewidth of practical lasers causes a spectrum with broaden lines as shown in figure 9.5a. From now on, we call this a comb spectrum.

The wider the laser linewidth, the wider all comb lines will be (see figure 9.5b), and the better the comb spectrum is converted into a flat spectrum with weak ripple. In a well-designed synthetic noise generator, this spectral ripple is very small. As a result, spectrum analyzers display a white spectrum, and are unable to distinct synthetic noise from random white noise.

Optimal delay time

The choice of this differential delay τ and of the modulation frequency f_m is inter-related. When $f_m = 1/(2\tau)$, then the two lightwave signals are 180° out of phase at when the two signals are combined. The lightwave frequencies oscillate symmetrically around a common mean value and the frequency sweep of the differential frequency becomes maximal. This is one of the reasons why this choice is an optimal choice.

Another reason to choose this value, is associated with suppression of parasitic IM modulation. The unwanted IM components in the lightwave signal interfere with the down-converted FM signal and infect the 'white' spectrum with equidistant spikes at a

multiple of the modulation frequency. When the lightwave signals in both branches are perfectly 180° out of phase, then full extinction of parasitic IM components will result. It will optimally suppress the odd harmonics of the IM signal in the photo current.

In conclusion:

$$\boxed{f_m = 1/(2\tau)} \quad \text{results in:} \quad \begin{cases} \text{maximum frequency sweep} \\ \text{optimal IM suppression} \end{cases}$$

We observed in our experimental setup that optimal IM suppression did not coincide with the above condition. It is plausible that this originates from small unbalance in the interferometer. In a practical setup, it is useful to de-tune the modulation frequency to improve the IM suppression.

Definitions

We summarize now some useful definitions, to simplify the description of synthetic noise generation.

- The *lightwave synthetic noise generator* is the lightwave device of figure 9.2, excluding the PIN-diode. Its output is the lightwave output of the interferometer. This PIN-diode can be any arbitrary diode, including the input diode of an optical receiver under test [801] or a lightwave noise-tee [802]. Therefore, this PIN diode is no part of the generator itself.
- The *synthetic noise spectrum* is the response spectrum of the photo current that flows through a PIN diode when illuminated by the lightwave noise source.
- The *homodyne spectrum* is the synthetic noise spectrum, when the periodical modulation is switched off.
- The *frame* of the synthetic noise spectrum is a discrete line spectrum that would have occurred when a perfect laser (with zero linewidth) is used. The elements of a frame are *frame lines* and have zero linewidth.
- The *comb* of the synthetic noise spectrum is the set of all individual spectra that are drawn in figure 9.5. The elements are *comb lines* and have a non-zero linewidth. The addition of all comb lines yields the synthetic noise spectrum.

9.1.2. Example of a practical synthetic noise generator

To demonstrate the feasibility and the performance of practical synthetic noise generators, we constructed an experimental setup with a DFB laser (1530 nm, 60 dB optical isolator), a fiber optic interferometer (about $\Delta L=10$ m differential delay, manual polarization adjustment) and a 10 MHz function generator. A block diagram of this setup is shown in figure 9.6.

The laser diode was modulated² in frequency (FM) by weak modulation of its bias current. This way of modulation, however, performs FM modulation as well as intensity modulation (IM). This IM effect is of minor importance since the process of FM to IM conversion in the interferometer largely dominates the parasitic IM modulation.

² In this experiment, triangular modulation and noise injection was used, as discussed in the succeeding sections.

The lightwave output signal was detected with a low noise optical receiver and a spectrum analyzer (HP8566b). The measured spectra were extracted by a computer to perform post processing of data. The detected dc photo-current was 100 μA .

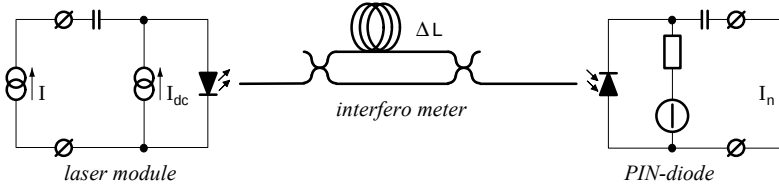


Fig 9.6 Block diagram of an implementation model of the lightwave synthetic noise generator. The current I represents the periodic modulation current, and I_{dc} represents the bias current.

Measurement and post processing. First, the synthetic noise spectrum was detected for various amplitudes of the 10 MHz signal. In addition, the system noise of the measurement system was detected by detecting the output noise when the illumination is removed from the optical receiver.

All these raw spectra are distorted by the unknown gain profile of the optical receiver and the unknown contribution of the noise by the measurement system. Therefore, the system noise was removed from the raw data, by subtraction of spectra on a power base. When U is the voltage that is detected with a selective voltmeter (spectrum analyzer), and when U_x and U_0 are the voltages that are associated with an illuminated and a non-illuminated receiver respectively, then the noise correction is performed with: $U'_x = \sqrt{(U_x^2 - U_0^2)}$.

Subsequently, the gain profile is reconstructed from the measured spectra by the assumption of a perfect white spectrum. To perform this, spikes, dips and weak fluctuations are stripped by proper interpolation and averaging techniques. When G is the smoothed function resulting from the spectrum originated from the deepest modulation, then the gain correction of all spectra is performed with $U'' = U' / G$.

Observations. Figure 9.7 demonstrates the post-processed noise spectra, for various modulation depths. The maximum modulation current, that was used in the experiment, is denoted with I_m . Its rms-value was about 12% of the laser bias current (60mA bias) to preserve linear FM modulation of the laser.

The spectral width is proportional to the maximum lightwave differential frequency, that is proportional to the amplitude of the modulation current I . The spectral width of the spectrum that originates from the modulation current $I_m/16$ is about 650 MHz. From this observation it is concluded that modulation with I_m will cause a spectral width that is 16 times wider, which is more than 10 GHz.

According to section 9.2.3 and 9.3.2, the rms-magnitude of the synthetic noise is independent of the modulation depth. As a result, the product of noise bandwidth and spectral intensity is invariant, which makes this density reciprocal to that current. Figure

9.7 shows that doubling the modulation depth results in 3 dB decrease of spectral intensity of the noise current.

The reconstructed spectra, that originate from modulation with the magnitudes $I_m/2$ and $I_m/4$, are perfectly white. From the precision that is associated with these spectra, it is concluded that the performed post processing is correct and accurate.

The reconstructed spectrum, that originates from modulation with the magnitude $I_m/8$, bears a slight deviation from white noise. The predicted spectral width is 1300 MHz for this current. From this weak deviation, and this prediction, it is concluded that approximately 50% of the synthetic noise spectrum is applicable for measurement purposes with white noise.

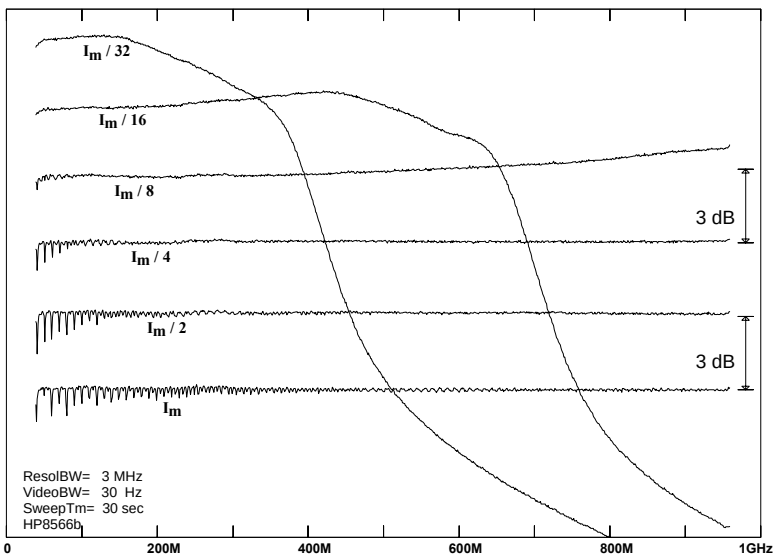


Fig 9.7 Measured output spectrum of a well-designed synthetic noise generator, for various modulation currents. All modulation currents are related to an arbitrary maximum value I_m . Its rms-value was about 12% of the laser bias current (60mA). Doubling the modulation depth resulted in twice the noise bandwidth and 3 dB decrease of spectral noise density. Note that triangular modulation and noise injection was performed as discussed in section 9.2.

Practical aspects. The small dips at the low end of the spectra in figure 9.7 originate from parasitic intensity modulation of the laser. The wider the spectrum, the deeper the modulation is required and the more the dips are pronounced in the spectrum. The use of a modern multi section laser, with separate inputs for IM and FM modulation, will probably nullify these side effects.

The maximum obtainable spectrum width is device dependent. A DFB laser, such as is used in this experiment, is not intended for frequency modulation. Its choice is therefore

not optimal. The use of modern multi section lasers will probably facilitate a spectrum width of 100 GHz or more.

9.1.3. Conclusions

In conclusion, we described the basic principles of a novel lightwave device, a synthetic noise generator. The instrument was original introduced as an alternative for bandwidth measurements, however we extended its applicability to lightwave noise measurements. We implemented an experimental setup, using a DFB laser diode, 10 MHz modulation and 10 m differential delay. We demonstrated that the experimental setup is usable up to 10 GHz. Multi section lasers may increase this band width to hundreds of GHz.

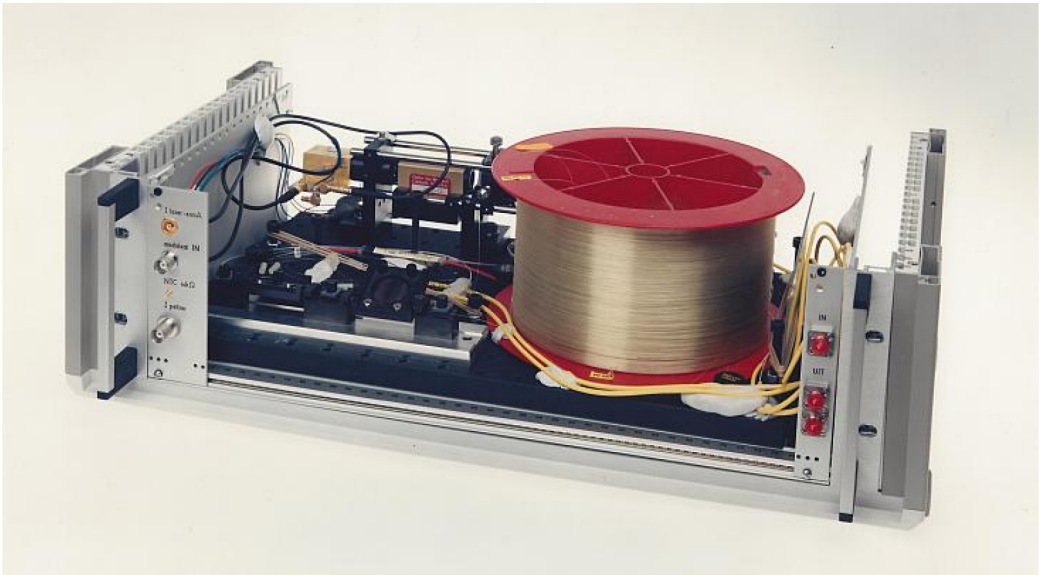


Fig 9.x Implementation of the lightwave synthetic noise generator, based on the diagram in fig 9.6.

9.2. Practical analysis of synthetic noise generation

The previous section was focused on a general understanding of synthetic noise generation, by discussing basic principles and the results from experiments and simulations. This section will focus on deeper understanding, and analyzes the principles of synthetic noise generation from a practical point of view. It uses the mathematical results from the theoretical analysis, as is described in section 9.3.

9.2.1. Analysis of the spectral envelope of the synthetic noise

The flat spectra in figure 9.7 have demonstrated that the spectral ripple in the synthetic noise spectrum can be made very small. So far, it was assumed that all ripple is caused by the distance between the individual comb lines in the spectrum. Unfortunately, this is not true. This subsection will demonstrate that the frame of the synthetic noise spectrum has a bumpy envelope, and that property dominates in the spectral ripple.

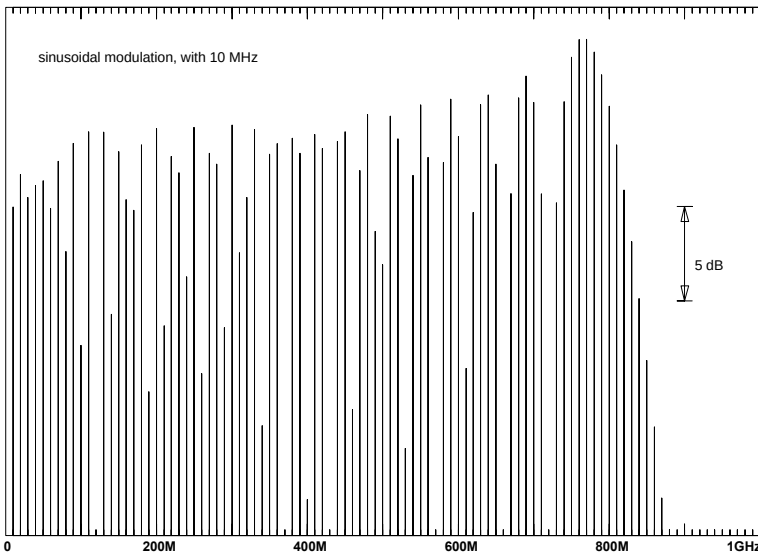


Fig 9.8 Simulated frame of the synthetic noise. This line spectrum originated from sinusoidal modulation. A fiber length of about 10 m is required to realize an optimal differential delay for 10 MHz modulation.

Figure 9.8 shows the simulation of a frame of the synthetic noise spectrum, in case of sinusoidal modulation. It is based on the theory of the succeeding section 9.3.3. The Bessel functions of figure 9.8 demonstrate that the frame envelope is quite rugged. Variations of more than 20 dB are observed for the intensity of succeeding frame lines. The simulation has further demonstrated that a weak change in modulation depth results in a significant variation of this envelope. The location and periodicity of the variations in the spectral envelope change with the modulation depth. As a result, the width of each comb line must span several frame lines to eliminate spectral ripple.

Another aspect that is observed from figure 9.8, is the increase with frequency of the mean envelope curve. Its deviation from a constant value prevents the synthetic noise spectrum from being usable as *white* noise over the full power bandwidth.

Wang et.al. [907] restricted their experiments to sinusoidal modulation only. They identified the problem of this deviation, derived an approximation for the mean envelope curve, however they did not find a solution. This work has resulted in a significant improvement to synthetic noise generation by using *triangular* modulation in stead of sinusoidal modulation.

Figure 9.9 shows the simulation of a frame of the synthetic noise spectrum under triangular modulation. It demonstrates that the mean envelope curve is significantly improved, compared to its equivalent in figure 9.8. The intensity variations of the individual frame lines are comparative in both figures, however the mean envelope curve in figure 9.9 is more constant over a wide frequency interval. Figure 9.10 is even more convincing.

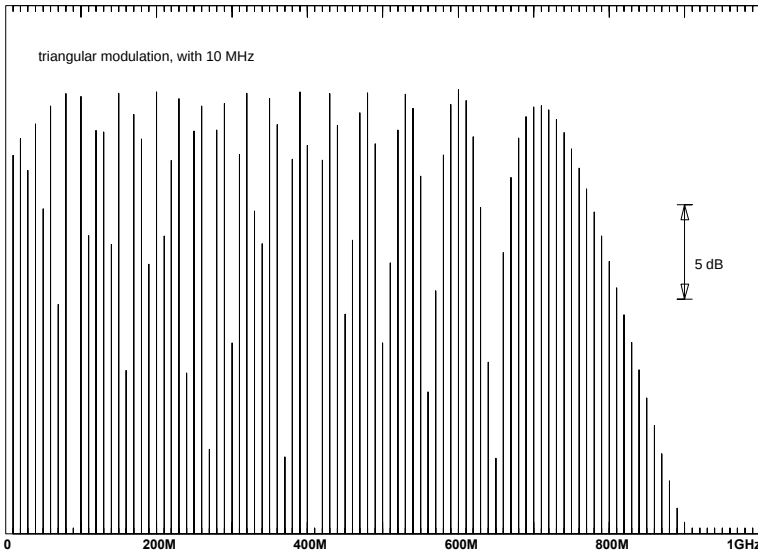


Fig 9.9 Simulated frame of the synthetic noise. This line spectrum originated from triangular modulation. A fiber length of about 10 m is required to realize an optimal differential delay for 10 MHz modulation.

As was stated before, the width of each comb line must span several frame lines to facilitate a smoothed synthetic noise spectrum. The more frame lines that are packed within the width of a comb line, the smaller the ripple will be in the synthetic noise spectrum. Further, the lower the modulation frequency f_m is chosen, the more tight the frame lines are packed. As a result, the lower the modulation frequency, the flatter the synthetic noise spectrum will be.

Figure 9.10 shows a simulation of the frame of a synthetic noise spectrum under triangular modulation, in which the modulation frequency is lower than in figure 9.9. It

demonstrates that the average frame envelope improves to that of the previous figure, and that the linewidth requirements of the laser are relaxed.

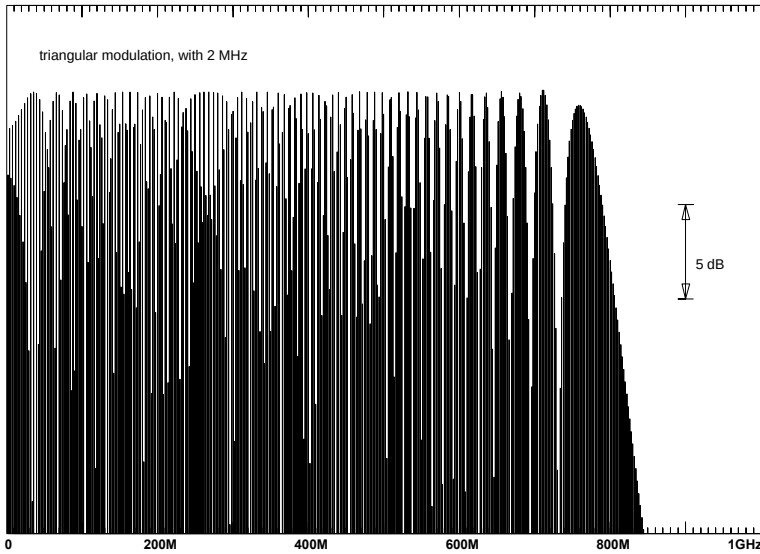


Fig 9.10 Simulated frame of the synthetic noise. This line spectrum originated from triangular. A fiber length of about 50 m is required to realize an optimal differential delay for 2 MHz modulation.

The draw-back of a decrease in modulation frequency f_m is the increase of differential length that the interferometer requires for an optimal match to f_m . A length of more than 10 kilometers is assumed to be unacceptable for a viable implementation, which sets a lower limit to the modulation frequency ($f_m > 10$ kHz). As a result, the lower the modulation frequency, the less viable the synthetic noise generator will be.

Figure 9.11 shows the measured synthetic noise spectrum of a deliberate poor-designed synthetic noise generator. The ripple in this synthetic noise spectrum demonstrates that the laser linewidth is too small for the chosen modulation frequency. The significant difference between ripple periodicity and modulation frequency demonstrates that the bumps in the frame envelope dominate in the spectral ripple.

As expected from the previous analysis, the experiment confirmed that the location and periodicity of the spectral ripple changes with the magnitude and the wave form of the modulation signal. Similar results were obtained from extended simulations that considered the envelope of the individual comb lines.

Another aspect that is confirmed by this experiment is that triangular modulation improves the mean spectrum. As was discussed before, the dips and spikes below 200 MHz result from the parasitic intensity modulation.

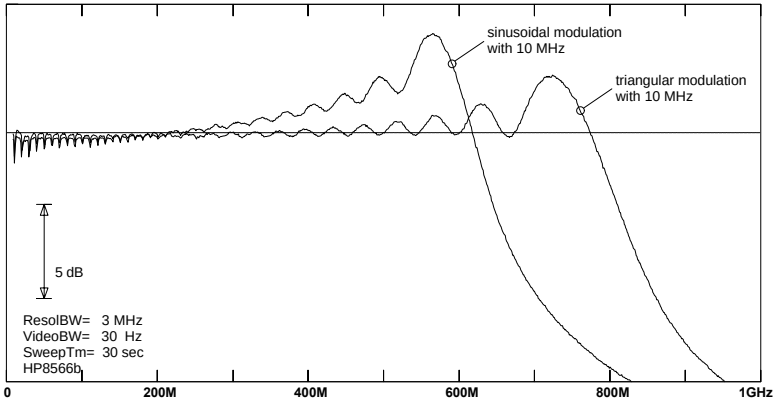


Fig 9.11 Measured output spectrum of the synthetic noise source, to demonstrate the ripple in the envelope of the FM line-spectrum. This ripple illustrates that the differential delay in the interferometer is too short with respect to the coherence length of the laser. Both spectra oscillate around an average that increases with frequency, however triangular modulation minimizes the inclination of this average. The rms-value of the periodical modulation current was approximately 0.8% of the laser bias current (60mA). Noise was not injected.

9.2.2. Spectral ripple reduction by noise injection

A ripple in the synthetic noise spectrum, that is as high as in figure 9.11, is unacceptable for noise measurement purposes. Wang et al. [907] kept this ripple as low as possible by choosing the differential delay τ in the interferometer significantly larger than the coherence time τ_c of the DFB laser. They implemented a delay with 1 km fiber, which facilitated that the optimal modulation frequency is relative low (100kHz). As a result, the comb lines are tighter packed than in figure 9.11, and the spectrum is more evenly smoothed by the laser linewidth (note the difference in envelope of the line spectra in figure 9.9 and 9.10).

The basic drawback of this approach is the delay line of more than 1 km fiber to fulfill these incoherence requirements. It will exclude the application of multi section lasers. These lasers are commonly optimized for minimum linewidth (maximum coherence length), and the use of these lasers would then require a delay line of even more kilometers.

We found a novel solution [901,902] to this discrepancy. We proposed broadening the linewidth of the individual comb lines instead of tighter packing. An artificial increase of the laser linewidth (equivalent with a decrease in coherence length) will broaden the width of the comb lines and will improve the smoothing of the synthetic noise spectrum. As a result, tighter packing of the comb lines becomes superfluous, and the application of relative high modulation frequencies and relative short delay lengths becomes adequate (e.g. 10 m and 10 MHz).

Basic principle

The idea of noise injection is based on the understanding that an additional FM modulation of the laser frequency with noise yields an artificial increase of its linewidth [909]. The random FM modulation results in lightwave frequency jittering, and this jitter manifests as a virtual increase of the laser linewidth.

The frequency jitter affects the laser frequency, the lightwave differential frequency and the comb lines in the down converted spectrum. The overall result is a synthetic noise spectrum with smoothed ripple.

Figure 9.12 shows the basic principle. A random noise signal is added to the periodic modulation signal, and the composed signal is used to modulate the laser frequency.

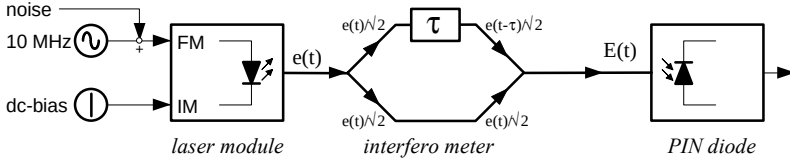


Fig 9.12 Block diagram of the synthetic noise generator, with additional noise injection. The electrical noise current is added to the periodical modulation current, to perform an additional random FM modulation. This results in an additional frequency jitter and performs an artificial increase of the laser linewidth.

Experimental verification

To demonstrate the impact of noise injection, we generated electrical noise. The spectral intensity was concentrated at relative low frequencies. Figure 9.13 shows the pink noise spectrum that was used for noise injection. This noise was superposed on the periodical modulation signal that generated figure 9.11. The injection level was set to approximately 7% (rms-value) of the laser bias current (60mA).

Noise injection with this relatively high injection level increases the laser linewidth significantly. Figure 9.14 demonstrates how the synthetic noise spectrum of figure 9.11 is smoothed by the proposed noise injection. The spectral ripple is vanished and the spectral width is equally increased with the linewidth.

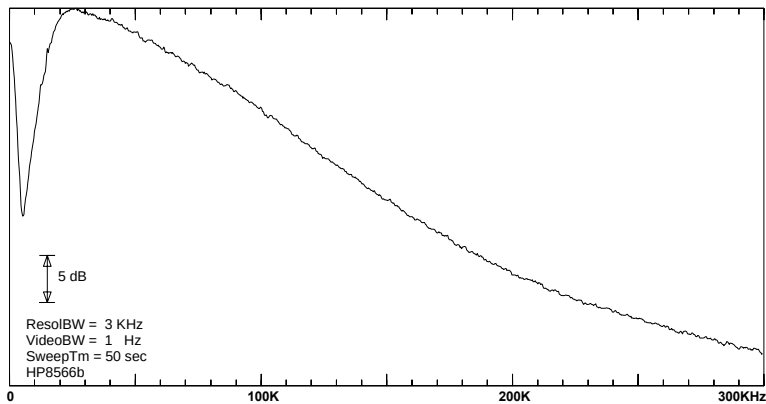


Fig 9.13 Measured spectrum of the pink noise that was injected in the synthetic noise generator.

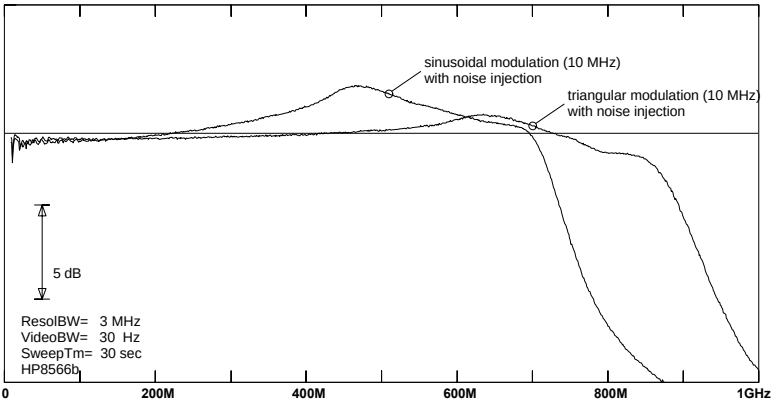


Fig 9.14 Measured output spectrum of the synthetic noise source, to demonstrate that noise injection smoothes the ripple in the synthetic noise spectrum.

The rms-value of the periodical modulation current was approximately 0.8% of the laser bias current (60mA). The injected noise level was approximately 7% of that bias current. This level was that high because the differential delay of the interfero meter was optimized to the 10 MHz modulation frequency and not for the spectrum in figure 9.13.

Spectrum of the injected noise

In the experimental verification we used pink noise, in stead of white noise, for noise injection. This is because we observed that the flatness of the laser linewidth curve improves when white noise injection is replaced by pink noise injection. To get a deeper understanding of this effect, we discuss the origin and shape of the laser linewidth in more detail.

Linewidth envelope: The intensity spectrum of a laser beam has essentially the well-known Lorentzian shape. When such a laser beam is detected with an appropriate delayed self-homodyne or self-heterodyne setup³, this shape is preserved in the responding spectrum. The width of the electrical response spectrum is twice the optical laser field (see Yariv [910, page 391]).

We will refer to that response spectrum as the *homodyne noise spectrum*. It equals the synthetic noise spectrum when the periodic modulation is switched off. As a result, the homodyne noise spectrum is Lorentzian, and its shape has the following well-known form: $S = S_0 / (1 + (f/\Delta f)^2) = S_0 / |1 + j f/\Delta f|^2$.

From a mathematical point of view, the homodyne noise spectrum is equivalent with the spectrum of a white noise source, filtered by a first order low-pass filter.

Linewidth model: The laser linewidth originates from spontaneous emission, and in addition from random electrical fluctuation such as shot noise. An adequate circuit, modeling this process, originates from a hypothetical zero linewidth laser and an intrinsic *white* noise source. The intrinsic noise source modulates the laser frequency and the resulting frequency fluctuations come through as linewidth (phase noise).

Shot noise, associated with the laser bias current, is one of the white noise currents that contributes to this intrinsic noise source. When the laser sensitivity for FM-modulation is known, then the rms-magnitude of this shot noise is indicative for the *minimum* linewidth that is expected for this laser. In practical lasers, the linewidth is significantly higher than expected from this assumption, which makes the equivalent noise source higher than the shot noise (see Henry [906] and Yariv [910, chapter 10]: linewidth enhancement factor α).

Linewidth envelope, using white noise injection: The natural linewidth of lasers virtually originates from FM modulation with internal white noise sources and has a Lorentzian shape. FM modulation of the laser frequency with *external* white noise sources yields an identical shape. Nevertheless, it will increase the laser linewidth, and thus the spectral width of the homodyne noise spectrum.

As a result, the homodyne noise spectrum with white noise injection has a first order fall-off, of which the -3 dB corner frequency marks the value of the laser linewidth. Above that corner frequency, the homodyne spectrum decreases with 6 dB per octave.

Linewidth envelope, using pink noise injection: Modulation of the laser frequency with *pink* noise will change both the width and the shape of the homodyne noise spectrum. We measured this modified shape of the homodyne noise spectrum using injection of pink noise with the spectrum of figure 9.13.

Figure 9.15 demonstrates these changes for various injection levels. It shows the synthetic noise spectrum when the periodical modulation is switched off (homodyne noise spectrum). At maximum injection level (7% of bias current), the spectrum is

³ Some authors use the word self-heterodyning while others use self-homodyning for the same subject. This thesis reserves the word self-heterodyning for setups that are similar to self-homodyne setups, but in which one beam is additionally shifted in frequency (e.g. 100 MHz with an acousto-optic modulator). In a delayed self-homodyne setup, two beams that originate from the same light source illuminate simultaneously a PIN diode. Since one beam is delayed, the two beams will incoherently mix in the photo diode to produce the differential spectrum. Both homodyne and heterodyne setups respond with similar spectra but with different center frequency. Self-heterodyning is more complex, but avoids dc-problems caused by optical fluctuations.

observed to be flat, up to 170 MHz. Above that corner frequency, the spectrum decreases with about 12 dB/oct. The more the injected noise is attenuated, the less the width and decay will be. When all noise injection is switched off, the homodyne spectrum becomes Lorentzian with a -3 dB corner frequency at about 17 MHz. As a result, the homodyne noise spectrum with pink noise injection has a second order fall-off. The homodyne noise spectra in figure 9.15 are similar to that of filtered white noise, that is filtered by second-order low-pass filters. Fair results were obtained with empirical curve fits, that originates from low-pass filters with one zero and two poles.

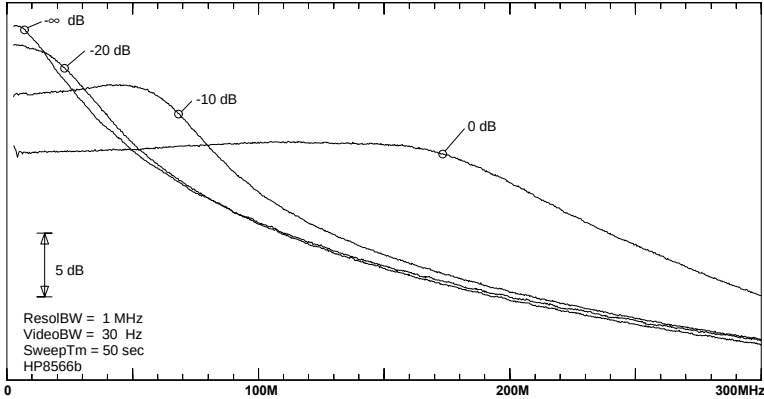


Fig 9.15 Measured synthetic noise spectrum in case the periodical modulation of the synthetic noise generator is switched off (homodyne noise spectrum). The level of the pink noise injection (see figure 9.13) is varied in 10 dB steps. These plots demonstrate that pink noise injection increases the homodyne spectral width and modifies its shape. Filtering of the injected noise has improved the flatness of the homodyne spectrum.

Benefits of pink noise injection: The use of pink noise injection, in stead of white noise injection, improves the smoothing 'efficiency' in the synthetic noise spectrum. This is because pink noise injection modifies the profile of the individual comb lines into a more 'rectangular' shape. As a result, the smoothing process between two succeeding comb lines is improved, and the spectral fall-off above the highest significant comb line is more abrupt. This makes the overall synthetic noise spectrum more 'rectangular' and the available noise power is more efficient concentrated within the user defined spectral width.

Simulation of homodyne spectra

The above mentioned experiments were restricted to variation of injection levels. The shape of the injected pink noise spectrum was not changed. Nevertheless, the spectral shape of this pink noise spectrum affects the homodyne spectrum too.

We did not experience with other noise spectra to inject, however we restricted these investigations to computer simulations. From these simulations, it was concluded that the shape of the injected noise spectrum is of minor importance, however the corner frequency of this low-pass profile is the most important figure of merit.

We simulated a delayed self-homodyne setup, using the time domain relations of section 9.3.1. The modulation frequency was set to $f_m = 10$ MHz and the associated optimal differential delay was set to $\tau = 1/(2 \cdot f_m)$. The injection noise signal was modeled by $15 \cdot 10^6$ Gaussian distributed random numbers. The pink noise spectrum was simulated, using a model of three cascaded first order filters with identical corner frequency. The homodyne spectrum was calculated using standard FFT techniques⁴. The injection magnitude of each simulation was adjusted to a value that facilitates equal output density at zero Hertz.

Figure 9.16 shows the simulated results for these $15 \cdot 10^6$ samples. The homodyne spectrum that originates from the 10 GHz filter (∞) has a Lorentzian shape, because it originates from injected noise that is nearly white. Filtering at 1 MHz changes the shape in a way that is more close to a rectangular profile. Further reduction of this corner frequency has no more than a minor effect to the shape.

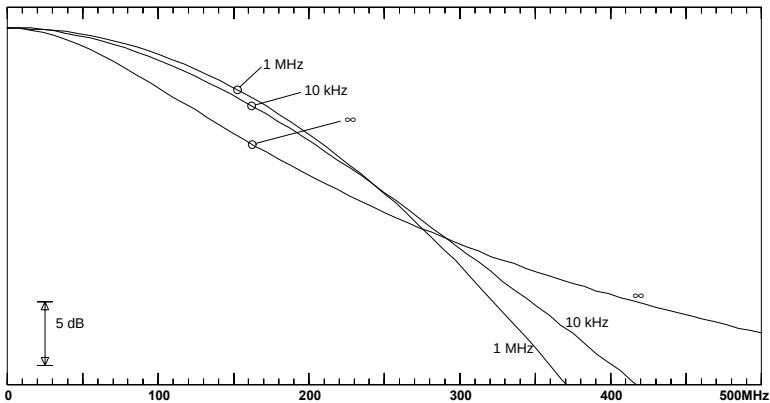


Fig 9.16 Simulated homodyne spectrum that results from pink noise injection, for various filter corner frequencies. The injected magnitude of each simulation was adjusted to facilitate equal output density at zero Hertz.

Similar simulations were performed at constant corner frequency of 1 MHz, however with reduced filter order. A noticeable change of shape was not observed. Other spectral adjustments to the injected noise will probably not result in a significant improvement of the homodyne spectrum, however forms an aspect of further research.

Optimal pink noise spectrum

From these computer simulations it is concluded that the filter corner frequency must be equal or lower than modulation frequency for which the synthetic noise generator is optimized. On the other hand, the lower the filter corner frequency is chosen, the more the two lightwave frequencies in the interferometer change with equal phase and, the

⁴ A single FFT calculation over $15 \cdot 10^6$ numbers was avoided using 450 individual FFT calculations over 32768 samples and combining these intermediate results.

lower the differential frequency will be. As a result, a higher injection level is required to preserve the spectral width of the comb lines in the synthetic noise spectrum. From this, it is plausible that optimal noise injection is associated with a filter corner frequency that is close to the modulation frequency. Note that this conclusion is *not* verified experimentally!

9.2.3. Power analysis of synthetic noise

A fair estimation of noise power can be obtained from simple dc current measurements. Furthermore a fair estimation of spectral noise density can be made from additional measurement on spectral bandwidth. These estimations provide simple means for rough exploratory measurement of absolute noise levels, without preceding calibration of the synthetic noise source. In this subsection we discuss their relation.

When a PIN photo diode is illuminated by the synthetic noise generator, then a photo current is generated. This photo current is the superposition of the (ac) synthetic noise and a dc-current. Two characteristic aspects of this photo current can be determined simply: mean value and spectral width. The mean value I_{n0} is obtained by measurement with a pica-ampere meter, and the spectral width B by measurement with a spectrum analyzer.

The mean value I_{n0} is a constant that depends on the chosen laser, its bias current and the attenuation in the optical path. Variation of the modulation depth will not affect this value. The spectral width B is a variable that is linear proportional to the modulation depth. When B is too wide for direct measurement with the used spectrum analyzer, then it must be derived from scaled measurements. In that case, extract B' from an equivalent measurement with reduced modulation, and scale B' to B with the used modulation reduction.

The synthetic noise originates from an incoherent mix of two lightwave signals. When the interference between these two signals varies from perfect extinction to perfect superposition, then the generated photo current varies from zero to a fixed maximum value. As a result, the photo current is a random current with rapid variations from zero to twice a mean value I_{n0} .

Under these conditions, the relation between the mean value and the rms-value of the ac current $i_n(t)$ can be evaluated simply (see section 9.3.2). Further, the relation between rms-value and spectral density S_i can be evaluated simply when the spectral width B is known. The product of S_i and B equals the square of the rms-value, according to the Parseval identity for spectra. For these quantities, we obtain:

I_{n0}	<i>= generated dc current</i>
$i_{rms} = I_{n0}/\sqrt{2}$	<i>= rms-magnitude of the noise</i>
$S_i(f) \approx (I_{n0}/\sqrt{2B})^2$	<i>= spectral intensity</i>
$\mu \stackrel{\text{def}}{=} (\sqrt{S_i})/I_{n0} \approx 1/\sqrt{(2 \cdot B)}$	<i>= noise-current ratio</i>

These expressions quantify how extremely efficient a synthetic noise generator is. Up to 70% of the generated dc-current is also available as noise. When this synthetic noise is

equally spread out over a full frequency band of $B=50$ GHz, and $I_{n0}=100\mu\text{A}$ dc-current flows through the photo diode, then the spectral noise density will be as high as $316\text{ pA}/\sqrt{\text{Hz}}$. This is 25 dB more than the thermal noise current of a 50Ω resistor. A dc-current of more than 300 mA would have been required to generate an equal spectral intensity by means of shot noise.

The definition of the noise-current ratio μ is similar to the well-known laser-RIN definition. In this example its value in [dB/Hz] is: $10 \cdot \log_{10}(\mu^2) = -110$ [dB/Hz]. This level is more than 40dB higher than typical RIN specifications for laser diodes. It illustrates that the ignorance of laser RIN effects does not vitiate our analysis.

From figure 9.7 it is observed that about 50% of the spectral width B is applicable for noise measurement, which is 25 GHz in this example.

For accurate scaling purposes, it is required that both I_{n0} and S_i are stable quantities. This is because ratio measurements require a precise variation of S_i , which is monitored by measuring the variation of the dc-current I_{n0} . Improvement of the individual components does not yield the required stability. Stabilization of the laser output power, the laser temperature and its modulation current is a commonly used measure. Further improvement of the individual components would require additional measures, such as stabilization of polarization drift in the interferometer, which makes it an inconvenient approach.

This work has resulted in novel techniques [901], to stabilize S_i by controlling the magnitude of the modulation current. This novelty is based on the understanding that an increase of the modulation depth will *increase* the spectral width B however will *decrease* the spectral density S_i . Figure 9.17 demonstrates the basic principle in which an additional control loop is used to stabilize I_{n0} simultaneously.

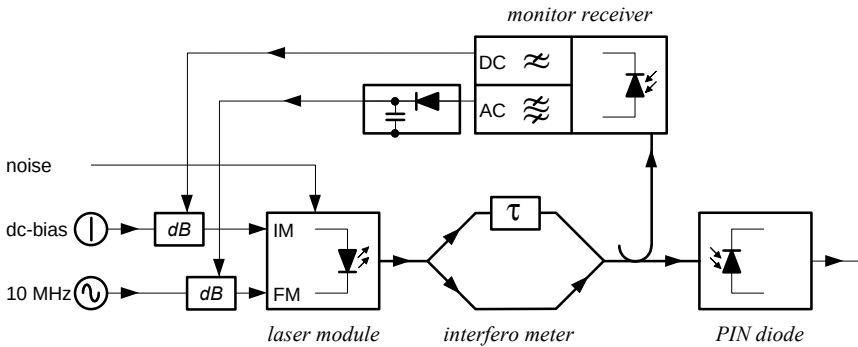


Fig 9.17 Dual loop method to stabilize the optical power as well as the intensity spectrum of the synthetic noise current. Both quantities are independently controlled.

The control loop monitors the output signal of the synthetic noise generator with an (internal) optical receiver. Two control signals are extracted simultaneously from this monitored signal, using the following methods:

- A low-pass filter extracts the dc-current from the monitor signal, and provides a first control signal. This signal is proportional to I_{n0} because it is proportional to the lightwave output power.
- A band-pass filter, cascaded with an envelope detector, provides a second control signal. This signal is proportional to the spectral density S_i if the pass-band of the filter is small, compared to the spectral width B . In practical implementations, the use of a high-pass filter will produce similar results, if the cut-off frequency of the monitor receiver is significantly lower than the spectral width B .

A first control loop decreases the laser bias current when the first control signal increases. This stabilizes the lightwave output power, and makes the setup insensitive to drift in laser power, laser temperature and coupling losses. Multiplication or addition of this signal with the laser bias current will fulfill these requirements.

A second control loop increases the magnitude of the (periodic) modulation signal when the second control signal increases. This stabilizes the spectral density S_i and makes it insensitive to polarization variations in the interferometer and other drift causes. It requires a controlled attenuator, a controlled amplifier or a multiplier.

Both loops are adequate to minimize drift in a synthetic noise generator, and make stabilization of the individual components superfluous.

9.2.4. Statistical analysis of synthetic noise

Natural noise sources commonly generate Gaussian distributed noise. Simple detection techniques, without true rms-detection, take advantage of this property for instance by estimating the noise rms-value from the detected average top level. The statistical distribution of synthetic noise is significantly different from a Gaussian distribution, and this results in detection errors when inadequate detection techniques are used. In this subsection we discuss the differences in statistical distribution.

The laser beam and the delayed laser beam are both modulated in frequency. The differential frequency, which is related to the differential phase $\Delta\alpha(t)$, is down-converted in the PIN diode, and forms the ac-part $i_n(t)$ of the photo current $I_n(t)$. This current has the following general form: $I_n(t) = I_{n0} \cdot (1 + \cos(\Delta\alpha(t))) = I_{n0} + i_n(t)$.

Since the phase of the laser beam is modulated with a periodic and a random signal, the differential phase is modulated too. In a well-designed synthetic noise generator, the differential delay and the modulation depth is large enough to make $(\Delta\alpha)_{rms} \gg \pi$. As a result, the photo current $I_n(t)$ changes rapidly and at random, and varies from zero to twice the mean photo current I_{n0} .

Figure 9.18 shows a simulated oscilloscope view of the photo current $I_n(t)$. The upper trace has a random appearance, however it has well-defined upper and lower magnitude limits. The lower trace is natural noise and has no limits at all. The difference in appearance demonstrates that the probability distribution of synthetic noise differs significantly from that of natural noise, which is usually Gaussian (normal)-distributed.

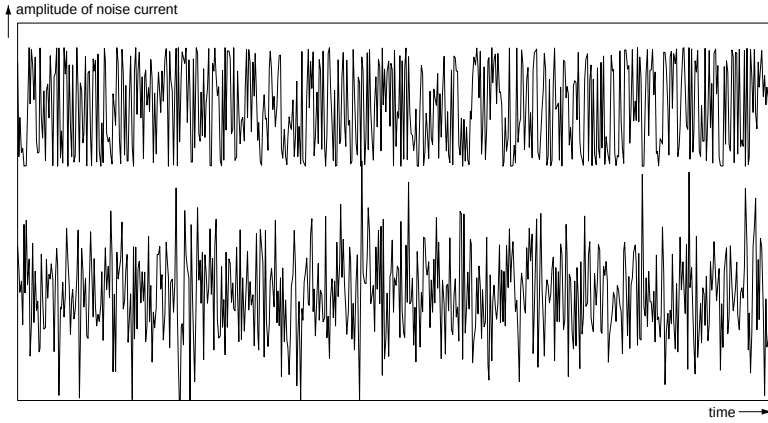


Fig 9.18 Simulated oscilloscope view of synthetic noise (upper trace) and natural noise (lower trace). The rms-values of both noisy signals are equal. Synthetic noise has a random appearance, however with well-defined upper and lower magnitude limits. This property does not hold for natural noise.

The upper trace is representative provided that the noise generator operates far above the incoherence threshold (see section 9.3.2). This holds for sinusoidal as well as triangular modulation; with or without noise injection.

When the rms-magnitude of $\Delta\alpha$ is large, then the magnitude distribution of $\Delta\alpha$ is irrelevant for the magnitude distribution of $\cos(\Delta\alpha)$. This is because the function $\cos(\Delta\alpha)$ wraps the results in 2π intervals, and causes the signal $\varphi = \arccos(\cos(\Delta\alpha))$ to be a random signal with near uniform-distribution. This makes the probability distribution proportional to $p(x) = (1/\pi) \sqrt{a^2 - x^2}$, for $|x/a| < 1$, a distribution that is named *trigonometric-distribution* in this text. Figure 9.19 shows the trigonometric-distribution of synthetic noise, in combination with the Gaussian-distribution of natural noise.

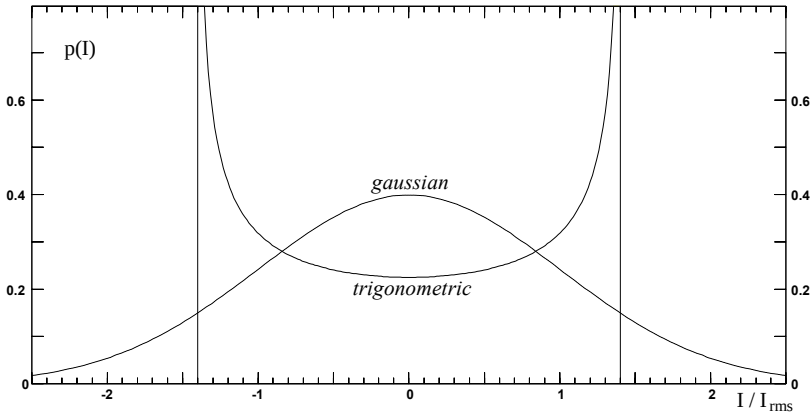


Fig 9.19 Probability distribution function of synthetic noise, normalized to its rms-value. The rms-value is 70% of the maximum ac-value ($=1/\sqrt{2}$). The function is compared to that of Gaussian distributed noise, with equivalent rms-value.

Figure 9.19 shows that the difference between trigonometric-distributed noise and Gaussian-distributed noise is significant. This aspect is important for precision calibration measurements on synthetic noise with calibrated noise standards.

Both a *linear* selective voltmeter and a *true-rms* detector are required for correct noise measurement. Nevertheless, many instruments that are specialized for this purpose do not fulfill both requirements. This effect limits their accuracy, which is discussed below.

rms-conversion uncertainty

Noise detection with an envelope detector, in stead of a true-rms detector, will cause erroneous results. Variation of the type of amplitude distribution, with conservation of spectral density, will cause a variation in read-out.

Some commonly used instruments, that are specialized for noise measurement, are sensitive to this type of variations. To quantify this sensitivity, we compared Gaussian-distributed noise of a standard noise source (HP346c) with trigonometric-distributed noise of the synthetic noise source, by using an HP8970a Noise Figure meter. The magnitude of the synthetic noise source was adjusted to facilitate two noise signals with equal spectral density. An internal envelope detector (instrument display) as well as an external true-rms detector (HP8481a power sensor, via 20 MHz IF-output) were used to make this comparison. Inequalities up to 0.2 dB were indicated by the rms-detector, for equal indication with the envelope detector.

As a result, the uncertainty in the rms-conversion factor decreases the overall measurement accuracy when the statistical properties of noise vary. This limits the performance of commonly used noise figure meters and spectrum analyzers

log-compression uncertainty

Noise detection with *logarithmic* selective voltmeters, in stead of *linear* selective voltmeters, will also cause erroneous results. This applies to most spectrum analyzers.

Logarithmic amplifiers do not preserve the statistical distribution of noisy signals, because the higher noise values are not amplified as much as the lower values. This causes a systematic error, due to the difference in instrument sensitivity between periodic and random signals, which requires a correction.

The noise marker of a spectrum analyzer is corrected for this effect, a correction that is optimized for Gaussian distributed noise. A log-correction of about 1.45 dB is required [804] when envelope detection and video filtering are used. Note, that this factor is irrespective to an additional rms-correction factor of 1.05 dB that is required [804] for noise detection with an envelope detector that is calibrated for rms-indication of sinus waves.

As a result, the uncertainty in the log-correction factor degrades the overall measurement accuracy when the statistical properties of noise vary. This is an additional performance limitation of commonly used spectrum analyzers, which does not hold for the above mentioned noise figure meter. In practical situations the total uncertainty of most spectrum analyzers is higher than the rms-conversion uncertainty and the log-compression uncertainty, which makes these uncertainties usually of minor importance.

9.2.5. Conclusions

In conclusion, we analyzed the spectral properties of the synthetic noise generator and demonstrated spectral ripple effects in case of insufficient incoherence. We discussed that a solution with long differential delay may fulfill the incoherence requirements, however excludes the use of multi section lasers.

We proposed a superior alternative, using injection of electrical noise, and analyzed various aspects of this injection. We demonstrated that injection with pink noise, filtered above the modulation frequency of the setup, provides optimal smoothing of the spectral ripple. Further we demonstrated that triangular modulation, instead of sinusoidal modulation, improves the usable portion of the noise bandwidth. Up to 50% of the spectral width B is applicable for noise measurements.

We demonstrated the high efficiency of synthetic noise generation. Up to 70% of the generated power is available as noise. The spectral width user defined, which facilitates the generation of very high noise levels. Synthetic noise is associated with very low (parasitic) dc-currents. This may be crucial for noise measurements on dedicated low-noise lightwave receivers.

For independent stabilization of optical power and spectral intensity we proposed a new dual loop method. We analyzed the statistical properties of synthetic noise, and discussed the implications of the difference between natural noise and synthetic noise.

9.3. Theoretical analysis of synthetic noise generation

The previous section analyzed many aspects of synthetic noise generation, from a practical point of view. It elucidated these aspects with the help of various simulation results, without discussing the underlying mathematics.

This section will focus on adequate mathematical models of synthetic noise generation, and provides the theory and formulae for the simulated results in the previous section.

9.3.1. Time domain analysis of the synthetic noise signal

A synthetic noise generator is essentially a lightwave oscillator, modulated in frequency. There are two FM modulation inputs, one for a periodic modulation signal $I(t)$ and another for a random signal $J(t)$. A third input for intensity modulation (IM) is required for controlling laser power.

Figure 9.20 shows an adequate calculation model of a synthetic noise generator, in which the FM modulation is represented by an integrator and a phase modulator. The light beams in both interferometer branches are assumed to be equally polarized.

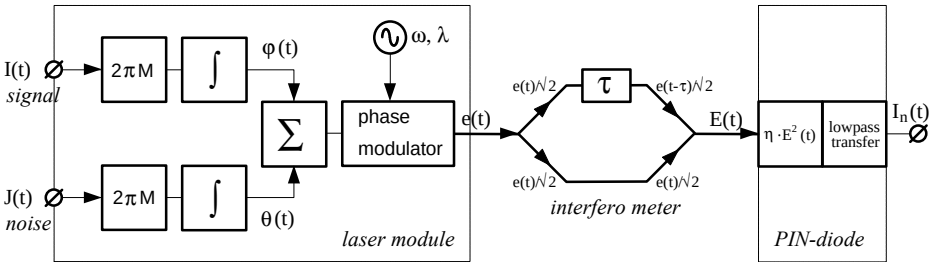


Fig 9.20 An adequate calculation model of the lightwave synthetic noise generator. $I(t)$ represents the periodical modulation current and $J(t)$ the combination of externally injected noise and internal parasitic noise. Both currents equally affect the laser frequency, however, this model isolates $I(t)$ from $J(t)$ to distinct $j(t)$ from $q(t)$.

Variation of the laser bias current yields a variation of lightwave intensity (IM) as well as the lightwave frequency (FM). Nevertheless, our analysis leaves all parasitic intensity modulation effects out of consideration. This is allowed because:

- The combination of interferometer and PIN diode makes the setup highly sensitive to lightwave FM signals and in practical situations the detected FM signal dominates the parasitic IM signal. This holds for our experimental setup with a DFB laser, and even more when multi section lasers are used. Multi section lasers have separated inputs for IM and FM modulation, while the modulation of DFB lasers is restricted to variation of their bias current.
- Laser intensity noise (RIN) is an example of parasitic IM effects. The calculation examples in section 9.2.3 and 9.3.2 illustrate that synthetic noise dominates the intensity noise (say 40 dB higher).

- Further, parasitic IM components in the photo diode response spectrum are suppressed when their frequency equals an odd multiple of $1/(2\tau)$. This is discussed in section 9.1 when selecting of the optimal delay time τ . When delay time and periodical modulation frequency are optimally matched, then all odd harmonics are suppressed.

As a result, IM modulation effects are of minor importance for the generation of synthetic noise.

The field strength of the modulated lightwave signal is $e(t)$, and this field is split, delayed and combined to the field strength $E(t)$. This signal is proportional to the differential field $e(t)-e(t-\tau)$, and generates a photo current $I_n(t)$ in a PIN diode. This current is the synthetic noise current.

In this subsection we will derive the relation between synthetic noise $I_n(t)$ and the two modulation signals $I(t)$ and $J(t)$. Let $\varphi(t)$ and $\theta(t)$ be the modulation phase at the input of the phase modulator, and let $\Delta\varphi(t,\tau)$ and $\Delta\theta(t,\tau)$ be their differential values with interval τ , then they are as follows related to the modulation signals $I(t)$ and $J(t)$:

$$\varphi(t) \stackrel{\text{def}}{=} 2\pi \cdot M \cdot \int_{-\infty}^t I(t) \cdot dt = (\text{periodic})$$

$$\theta(t) \stackrel{\text{def}}{=} 2\pi \cdot M \cdot \int_{-\infty}^t J(t) \cdot dt = (\text{random})$$

$$\Delta\varphi(t,\tau) \stackrel{\text{def}}{=} \varphi(t)-\varphi(t-\tau) = 2\pi \cdot M \cdot \int_{t-\tau}^t I(t) \cdot dt = 2\pi \cdot M \cdot \int_{-\infty}^t \{I(t)-I(t+\tau)\} \cdot dt$$

$$\Delta\theta(t,\tau) \stackrel{\text{def}}{=} \theta(t)-\theta(t-\tau) = 2\pi \cdot M \cdot \int_{t-\tau}^t J(t) \cdot dt = 2\pi \cdot M \cdot \int_{-\infty}^t \{J(t)-J(t+\tau)\} \cdot dt$$

Let ω_0 be a constant lightwave frequency ($\omega_0=2\pi \cdot c/\lambda$) that is generated in the ideal lightwave oscillator. Let ω be the momentaneous frequency of the laser output signal. This frequency value ω is fluctuating due to spontaneous emission (phase noise) and thermal drift. The model of figure 9.20 represents all these fluctuations by an integral part of the noise source $J(t)$. This representation facilitates the use of an initial *lightwave* phase ϕ_0 of the unmodulated laser at $t=0$.

The lightwave signal is divided into two beams in the interfero meter. Let ΔL be the difference in fiber length between the two branches of the interfero meter, then the associated delay ($\tau=\Delta L/c$) results in a phase difference $\Delta\alpha(t,\tau)$ between the two beams when are combined. This differential phase is as follows related to the previous quantities:

$$\alpha_r \stackrel{\text{def}}{=} \phi_0 + \omega_0 \cdot t + \varphi(t) + \theta(t) = \text{reference phase}$$

$$\alpha_d \stackrel{\text{def}}{=} \phi_0 + \omega_0 \cdot (t-\tau) + \varphi(t-\tau) + \theta(t-\tau) = \text{delayed phase}$$

$$\alpha_\tau \stackrel{\text{def}}{=} (\omega_0 \cdot \tau \oplus 2\pi) = 2\pi \cdot (\Delta L/\lambda \oplus 1) = \text{offset phase}$$

$$\Delta\alpha \stackrel{\text{def}}{=} \alpha_\tau + \Delta\varphi(t,\tau) + \Delta\theta(t,\tau) = (\alpha_r - \alpha_d) + k \cdot 2\pi = \text{differential phase}$$

In these expressions, the operation $(x \oplus y)$ denotes the modulo operation that returns the remainder of the division of (x/y) with an integer result. The current $I_n(t)$, that responds in the photo diode from illumination with the lightwave field strength $E(t)$, is now derived as follows:

$$\begin{aligned} e(t) &= E_0 \cdot \cos(\phi_0 + \omega_0 \cdot t + \varphi(t) + \theta(t)) \\ &= E_0 \cdot \cos(\alpha_r) \end{aligned}$$

$$\begin{aligned} E^2(t) &= (\frac{1}{2} \cdot e(t) + \frac{1}{2} \cdot e(t-\tau))^2 \\ &= (\frac{1}{2} \cdot E_0)^2 \cdot \{\cos(\alpha_r) + \cos(\alpha_d)\}^2 \\ &= \frac{1}{4} \cdot E_0^2 \cdot \{\cos^2(\alpha_r) + 2 \cdot \cos(\alpha_r) \cdot \cos(\alpha_d) + \cos^2(\alpha_d)\} \\ &= \frac{1}{4} \cdot E_0^2 \cdot \{\frac{1}{2} + \frac{1}{2} \cdot \cos(2\alpha_r) + \cos(\alpha_r + \alpha_d) + \cos(\alpha_r - \alpha_d) + \frac{1}{2} + \frac{1}{2} \cdot \cos(2\alpha_d)\} \\ &= \frac{1}{4} \cdot E_0^2 \cdot \{1 + \cos(\alpha_r - \alpha_d) + \cos(\alpha_r + \alpha_d) + \frac{1}{2} \cdot \cos(2\alpha_r) + \frac{1}{2} \cdot \cos(2\alpha_d)\} \end{aligned}$$

$$\begin{aligned} I_n(t) &= (\frac{1}{4} \cdot \eta E_0^2) \cdot \{1 + \cos(\alpha_r - \alpha_d)\} + \text{Currents at Lightwave Frequencies} \\ &= (\frac{1}{4} \cdot \eta E_0^2) \cdot \{1 + \cos(\omega_0 \cdot \tau + \Delta\varphi(t, \tau) + \Delta\theta(t, \tau))\} \\ &= (\frac{1}{4} \cdot \eta E_0^2) \cdot \{1 + \cos(\omega_0 \cdot \tau \oplus 2\pi) + \Delta\varphi(t, \tau) + \Delta\theta(t, \tau)\} \\ &= (\frac{1}{4} \cdot \eta E_0^2) \cdot \{1 + \cos(\alpha_r + \Delta\varphi(t, \tau) + \Delta\theta(t, \tau))\} \\ &= (\frac{1}{4} \cdot \eta E_0^2) \cdot \{1 + \cos(\Delta\alpha(t, \tau))\} \end{aligned}$$

$I_n(t) \stackrel{\text{def}}{=} I_{n0} \cdot (\cos(\Delta\alpha(t, \tau)) + 1)$ $i_n(t) \stackrel{\text{def}}{=} I_{n0} \cdot \cos(\Delta\alpha(t, \tau))$ $I_{n0} \stackrel{\text{def}}{=} (\frac{1}{4} \cdot \eta E_0^2)$ $\Delta\alpha(t, \tau) = (\omega_0 \cdot \tau \oplus 2\pi) + 2\pi \cdot M \cdot \int_{t-\tau}^t (I(t) + J(t)) \cdot dt$

These equations demonstrate that the magnitude of the output signal spans a restricted range, because of the cosine term. The minimum value is zero and the maximum is $2 \cdot I_{n0}$. We used these expressions to simulate the plots in figure 9.18. The simulations were performed with MatLab, using the following steps:

- 1): $J(t) = \text{array with Gaussian distributed random numbers}$
- 2): $\Delta\theta(t, \tau) = 2\pi \cdot M \cdot \int_{-\infty}^t \{J(t) - J(t+\tau)\} \cdot dt$
- 3): $I_n(t) \stackrel{\text{def}}{=} I_{n0} \cdot (\cos(\Delta\theta(t, \tau)) + 1)$

Without any modulation, $I(t)=J(t)=0$, or without any differential delay, $\tau=0$, the differential phase $\Delta\alpha$ is zero, which causes the photo current to be maximal.

9.3.2. Incoherence analysis

The interference between the two light beams in the synthetic noise generator causes a variation in the photo current $I_n(t)$, and the associated ac-current $i_n(t) = I_{n0} \cdot \cos(\Delta\alpha)$ represents the synthetic noise. Periodic modulation will result in deterministic variations of $I_n(t)$, and this situation is referred as a coherent interference. Superposition of a random modulation signal will change this into an incoherent interference, which results in random variation of the photo current.

Incoherent interference is essential for synthetic noise generation. Therefore this subsection derives what values are required for the random modulation $J(t)$ and the differential delay time τ , to perform adequate incoherent interference. It will be demonstrated that this demand is equivalent with the restriction that the differential delay τ must exceed the coherence time τ_c of the laser.

Definition of incoherence threshold

The differential phase $\Delta\alpha$ is the superposition of periodic variations $\Delta\varphi(t,\tau)$, random fluctuations $\Delta\theta(t,\tau)$, and an offset value. The magnitude of these random fluctuations are indicative for the deviation from pure coherent interference, and we define:

<i>Incoherence</i>	$\stackrel{\text{def}}{=} \Delta\theta_{\text{rms}}$
<i>Incoherence threshold:</i>	$\Delta\theta_{\text{rms}} \stackrel{\text{def}}{=} \pi$

The incoherence increases with the depth of random modulation, however increases with the differential delay τ . Figure 9.21 illustrates the relation between a logarithmic increase of incoherence and the increase of the random fluctuations in the photo current $I_n(t)$.

The deeper the laser is random modulated, the higher the incoherence will be and the more ac-current is generated. When the incoherence equals the value π , then $\cos(\Delta\theta)$ will fluctuate over the full available span. This special value will be referred in this text as the *incoherence threshold*.

Above incoherence threshold, the interference is wrapped in 2π intervals and the current fluctuations span the full available range. The interference between the two lightwave signals will then fluctuate from perfect extinction to perfect superposition.

Figure 9.21 illustrates that the photo current ranges from zero to $2 \cdot I_{n0}$. Far below the incoherence threshold, the mean value I_{n0} of the photo current $I_n(t)$ approaches this maximum value, however far above the incoherence threshold the mean value has been reduced to half the maximum photo current.

Further increase of the differential phase will not increase the ac noise magnitude, because it cannot exceed the value I_{n0} . This makes the total available noise power *independent* of the modulation depth. From this, it is concluded that a well-designed synthetic noise generator must always operate far above the incoherence threshold.

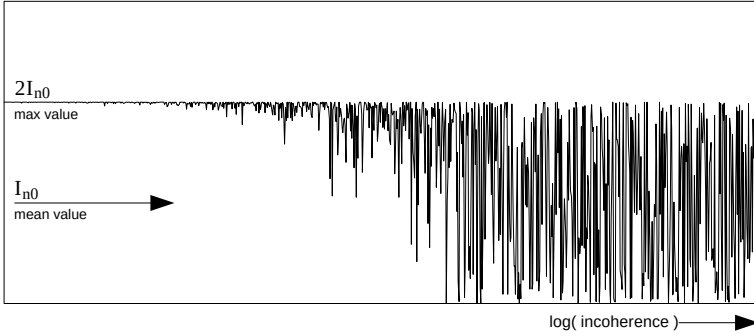


Fig 9.21 Simulated oscilloscope view of the synthetic noise current $I_n(t)$, in case the random modulation $J(t)$ sweeps from zero to a maximum value. Far above incoherence threshold, the output noise spans the full range, and its mean value I_{n0} has been reduced to half the maximum value.

Incoherent power analysis

Far above incoherence threshold, the mean of $\cos(\Delta\theta)$ approaches to zero, and facilitates a simple relation between rms-magnitude of the ac noise current and the mean value of the dc current:

$$\begin{aligned} (i_{\text{rms}})^2 &= \langle i_n^2(t) \rangle_\infty = I_{n0}^2 \cdot \langle \cos^2(\Delta\theta) \rangle_\infty = I_{n0}^2 \cdot \langle \frac{1}{2} + \frac{1}{2} \cos(2\Delta\theta) \rangle_\infty \\ &= I_{n0}^2 \cdot \langle \frac{1}{2} \rangle_\infty \end{aligned}$$

$$\boxed{i_{\text{rms}} = I_{n0}/\sqrt{2} = \text{rms magnitude of the noise}} \quad \text{far above the incoherence threshold}$$

In a well-designed noise generator, the synthetic noise is equally spread out over a wide frequency band B . This width is defined by the frequency modulation depth of $I(t)$. Since the Parseval identity relates the rms-value to the single-sided spectral density $S_i(f)$ of the noise, we obtain for white noise:

$$(i_{\text{rms}})^2 = \langle i_n^2(t) \rangle_\infty = \int_0^{+\infty} S_i(f) \cdot df \approx S_i(0) \cdot B$$

$$\boxed{S_i(f) \approx (I_{n0}/\sqrt{2B})^2 = \text{spectral power density}} \quad \text{far above the incoherence threshold}$$

The ratio between spectral density and dc photo current is an important characteristic figure of the synthetic noise source, since it enables the specification of noise at multiple noise levels. It is similar to the definition of laser RIN that is commonly specified in dB/Hz using $10 \cdot \log_{10}(\mu^2)$:

$$\boxed{\mu \stackrel{\text{def}}{=} (\sqrt{S_i})/I_{n0} \approx 1/\sqrt{(2 \cdot B)} = \text{noise-current ratio}} \quad \text{far above the incoherence threshold}$$

These formulae demonstrate the validation of conclusions in section 9.2.3, and the observation in section 9.1.2 that the intensity spectrum is reciprocal with the modulation current.

Incoherence sensitivity to differential delay and noise injection

The incoherence $\Delta\theta_{\text{rms}}$ increases with the magnitude of the injected noise current $J(t)$ and with the differential delay time τ . Since this is an important aspect of a well-designed synthetic noise generator we calculate the relation between rms-magnitude of the differential phase fluctuation $\Delta\theta$ and of the injected noise current $J(t)$.

Using the Parseval identities, to obtain the rms-magnitude of $\Delta\theta(t, \tau)$ from its single sided intensity spectrum $S_{\Delta\theta}(f)$, we get:

$$(\Delta\theta_{\text{rms}})^2 = \langle |\Delta\theta(t, \tau)|^2 \rangle_{\infty} = \int_0^{+\infty} S_{\Delta\theta}(f) \cdot df = \int_0^{+\infty} \mathcal{S}\{\Delta\theta(t, \tau)\} \cdot df$$

$$\text{using } \Delta\theta(t, \tau) = 2\pi \cdot M \cdot \int_{-\infty}^t \{J(t) - J(t+\tau)\} \cdot dt \quad \text{we obtain:}$$

$$(\Delta\theta_{\text{rms}})^2 = (2\pi \cdot M)^2 \cdot \int_0^{+\infty} \mathcal{S}\left\{ \int_{-\infty}^t \{J(t) - J(t+\tau)\} \cdot dt \right\} \cdot df$$

$$(\Delta\theta_{\text{rms}})^2 = (2\pi \cdot M)^2 \cdot \int_0^{+\infty} \left| \frac{1}{j\omega} \right|^2 \cdot \mathcal{S}\{J(t) - J(t+\tau)\} \cdot df$$

$$(\Delta\theta_{\text{rms}})^2 = (2\pi \cdot M)^2 \cdot \int_0^{+\infty} \left| \frac{1 - \exp(j\omega\tau)}{j\omega} \right|^2 \cdot \mathcal{S}\{J(t)\} \cdot df$$

After some trigonometric manipulations and the use of the shortcut: $\text{sinc}(x) = \sin(\pi x) / (\pi x)$, we obtain:

$$(\Delta\theta_{\text{rms}})^2 = (2\pi \cdot M)^2 \cdot \int_0^{+\infty} \tau^2 \cdot \text{sinc}^2(f\tau) \cdot \mathcal{S}\{J(t)\} \cdot df$$

$\Delta\theta_{\text{rms}} = (2\pi \cdot M \cdot \tau) \sqrt{\int_0^{+\infty} \text{sinc}^2(f\tau) \cdot S_j(f) \cdot df} = \text{Incoherence}$

This expression demonstrates that the incoherence increases with the rms-magnitude of $J(t)$ and the differential delay time τ . Further simplification of the above expression is feasible for two extreme situations: in case of pink noise and of white noise injection.

- Pink noise is defined as noise of which the intensity spectrum S_j decreases with increasing frequency. This facilitates the approximation $\text{sinc}^2(f\tau) \approx 1$ for all frequencies below $f = 1/(2\tau)$, and the approximation of $S_j = 0$ above this limit. Note, that the corner frequency $1/(2\tau)$ equals the modulation frequency f_m , in a well-designed synthetic noise source.
- White noise is defined as noise of which the intensity spectrum S_j is frequency independent over a sufficient wide frequency interval. This makes $S_j(f) = S_j(0)$ for all frequencies where $\text{sinc}^2(f\tau)$ is significant.

Shot noise is an example of white noise, and is indicative for the minimum injected noise. It originates from the laser bias current I_{bias} , and has $S = 2q_0 \cdot I_{\text{bias}}$ as spectral density. Note that a practical semiconductor laser suffers from spontaneous emission, which may increase this minimum value significantly.

Filtered noise is an example of pink noise, if the corner frequency of the low-pass filter is significantly lower than the modulation frequency.

In the two restricted situations, the expression for the incoherence reduces to:

$$\left[\begin{aligned} [\Delta\theta_{\text{rms}}]_{\text{pink}} &\approx (2\pi \cdot M \cdot \tau) \sqrt{\text{sinc}^2(0) \cdot \int_0^{+\infty} S_j(f) \cdot df} \\ [\Delta\theta_{\text{rms}}]_{\text{white}} &\approx (2\pi \cdot M \cdot \tau) \sqrt{S_j(0) \cdot \int_0^{+\infty} \text{sinc}^2(f\tau) \cdot df} \end{aligned} \right]$$

Using formula 15.36 from the Mathematical handbook [911] for the integral of a sinc-function, we obtain:

$$\left[\begin{aligned} [\Delta\theta_{\text{rms}}]_{\text{pink}} &\approx (2\pi \cdot M \cdot \tau) \cdot \sqrt{1 \cdot \langle |J(t)|^2 \rangle_{\infty}} \\ [\Delta\theta_{\text{rms}}]_{\text{white}} &\approx (2\pi \cdot M \cdot \tau) \cdot \sqrt{S_j(0) / (2 \cdot \tau)} \end{aligned} \right]$$

In the special case that all white noise originates from the shot noise of a dc-current I_{bias} , we may substitute:

$$S_j(0) = 2 q_0 I_{\text{bias}}$$

pink noise:	$\Delta\theta_{\text{rms}} \approx 2\pi \cdot M \cdot \tau \cdot \frac{J_{\text{rms}}}{\sqrt{2}}$
white noise:	$\Delta\theta_{\text{rms}} \approx 2\pi \cdot M \cdot \sqrt{\tau \cdot S_{j0}}$
shot noise:	$\Delta\theta_{\text{rms}} \approx 2\pi \cdot M \cdot \sqrt{\tau \cdot q_0 \cdot I_{\text{bias}}}$

These expressions demonstrate that the incoherence is linear proportional to the magnitude of the injected noise $J(t)$. The sensitivity to noise injection is affected by the differential delay time τ of the interferometer.

Incoherence condition

In a well-designed synthetic noise generator, the modulation frequency f_m is optimally matched to the differential delay τ of the interferometer, caused by differential length ΔL . This means that $f_m = 1/(2\tau) = c/(2 \cdot \Delta L)$, in which c is speed of light in the used fiber. Operation above incoherence threshold set additional prior conditions to the injected noise $J(t)$ and the modulation frequency f_m of the modulation signal $I(t)$.

The relation between incoherence, the random modulation magnitude and the differential delay is described before. As a result, operation above incoherence threshold requires one of the following conditions:

LF pink noise:	$f_m = 1/(2\tau) = c/(2 \cdot \Delta L) < M \cdot J_{\text{rms}}$
white noise:	$f_m = 1/(2\tau) = c/(2 \cdot \Delta L) < M^2 \cdot S_j(0)$
shot noise:	$f_m = 1/(2\tau) = c/(2 \cdot \Delta L) < 2M^2 \cdot q_0 \cdot I_{\text{bias}}$

These expressions demonstrate that compensation of a shortage in differential delay requires an abundance in noise modulation to perform incoherent operation.

A rough estimation of the laser linewidth is $\Delta f \approx 2 \cdot M \cdot J_{\text{rms}}$. Using this approximation for substitution in the incoherence condition for white noise modulation, then this condition reduces to $f_m = (1/2\tau) < (\Delta f)/2$. Since the coherence time t_c is related to this width by $\tau_c = \Delta f$, the incoherence condition is equivalent with the condition that the differential delay must exceed the coherence length.

This validates one of the conclusions of the Wang experiment [907] that the delay τ in the homodyne set-up should be larger than the coherence time of the laser diode.

practical calculation example

The shot noise condition is mainly a hypothetical condition since other random effects, such as spontaneous emission, dominate in practical semiconductor lasers. This can be concluded from the RIN (relative intensity noise) that is specified for commercially available laser diodes.

RIN is defined for a specific laser bias current I_{bias} when the laser under test illuminates a photo diode. Let S_i be the detected intensity spectrum of a laser in *excess* to the receiver noise and the shot noise of the generated photo current I_{photo} . Now RIN is defined as the ratio $\mu^2 = S_i/(I_{\text{dc}})^2$, and is commonly specified in [dB/Hz] using $10 \cdot \log_{10}(\mu^2)$.

Further, let R be the current responsivity of the setup, being defined as $R = I_{\text{photo}}/I_{\text{bias}}$. The relation between the laser intensity noise S_i (at the photo diode side) and the equivalent noise S_j (at the laser side) 'generating' this intensity noise is as follows:

$$\mu^2 = \frac{S_i}{(I_{\text{photo}})^2} = \frac{R^2 \cdot S_j}{(R \cdot I_{\text{bias}})^2} = \frac{S_j}{(I_{\text{bias}})^2} \stackrel{\text{def}}{=} \frac{\alpha \cdot (2q_0 \cdot I_{\text{bias}})}{(I_{\text{bias}})^2} = \alpha \cdot \left(\frac{2q_0}{I_{\text{bias}}} \right)$$

In these expressions is α an enhancement factor that indicates how much the spontaneous emission dominates the shot noise of the laser bias current.

In the hypothetical case that $\alpha=1$, than the RIN level would have been -172 dB/Hz for 50 mA laser bias current. Usually, the specified RIN level is 20 dB higher which yields that $\alpha=100$.

To quantify the incoherence condition for the above laser example, assume that the FM sensitivity of the laser module is $M = 300$ MHz/mA and a bias current of $I_{\text{bias}}=50$ mA. On one hand, this generator requires at least $J_{\text{rms}}=33\mu\text{A}$ of pink noise when a modulation frequency of $f_m=10$ MHz ($\Delta L \approx 10$ m) is optimal and when all noise is injected as pink noise. On the other hand, this generator requires at most an optimal modulation frequency of $f_m=140$ kHz ($\Delta L \approx 700$ m) when all noise originates from laser intensity noise ($\alpha=100$ times the shot noise of 50 mA bias current).

9.3.3. Spectral analysis far above incoherence threshold

The time domain analysis of subsection 9.3.1 requires detailed information on the phase of the lightwave signal, with respect to $t=0$. This phase information is available for the periodic modulation signal $I(t)$, however the noise signal $J(t)$ is random and the internal part of $J(t)$ is not available for measurement.

A well-designed synthetic noise generator operates above incoherence threshold, and one of the consequences is that the statistical properties and the intensity spectrum of the synthetic noise are not affected by an arbitrary phase offset of the lightwave signal.

This subsection demonstrates that operation above incoherence threshold yields a significant simplification of spectral calculus. Further it applies this simplification to spectral analysis of the full synthetic spectrum.

Mean values above incoherence threshold

When $\Delta\theta(t)$ is a random signal with zero mean, and ϕ_0 is a constant phase offset, then the modulated signal $\cos(\Delta\theta(t)+\phi_0)$ is random too. This signal has a mean value that depends on the phase offset term ϕ_0 . This dependency is not linear and exists for small modulation signals only. Far above or far below the incoherence threshold this dependency is approximated by:

$$\begin{aligned} \langle \cos(\Delta\theta(t)+\phi_0) \rangle_\infty &\approx \cos(\phi_0) && \text{for near coherent situations, when } \Delta\theta_{\text{rms}} \ll \pi \\ \langle \cos(\Delta\theta(t)+\phi_0) \rangle_\infty &\approx 0 && \text{for incoherent situations, when } \Delta\theta_{\text{rms}} \gg \pi \end{aligned}$$

Operation above incoherence threshold was defined in section 9.3.2 and means that $\Delta\theta_{\text{rms}} > \pi$. As a result, the differential phase $\Delta\theta$ fluctuates over more than a 2π span. Therefore, the constant phase offset ϕ_0 will hardly affect the statistical (and spectral) properties of the modulated signal. For an arbitrary phase offset ϕ_0 , an arbitrary delay time t_0 , and arbitrary random values $\Delta\theta(t)$ with *sufficient* magnitude (far above the incoherence threshold), this results in:

$$\begin{aligned} \langle \sin(\Delta\theta(t)) \rangle_\infty &= 0 && \langle \cos(\Delta\theta(t)) \rangle_\infty = 0 \\ \langle \sin(\Delta\theta(t)+\phi_0) \rangle_\infty &= 0 && \langle \cos(\Delta\theta(t)+\phi_0) \rangle_\infty = 0 \\ \langle \sin(\Delta\theta(t)+\Delta\theta(t+t_0)) \rangle_\infty &= 0 && \langle \cos(\Delta\theta(t)+\Delta\theta(t+t_0)) \rangle_\infty = 0 \\ \langle \sin(\Delta\theta(t)-\Delta\theta(t+t_0)) \rangle_\infty &= 0 && \langle \cos(\Delta\theta(t)-\Delta\theta(t+t_0)) \rangle_\infty = \delta \\ &&& \text{if } |t_0| \gg 0 \text{ then } \delta \rightarrow 0 \\ &&& \text{if } |t_0| \rightarrow 0 \text{ then } \delta \rightarrow 1 \end{aligned}$$

As a result, most of these examples have zero mean, with an exception for the cosine differential case. It approaches to the value 1, in the limit that the differential delay approaches to zero.

Intensity spectra far above incoherence threshold

The calculation of random modulated FM spectra is simplified when the modulation exceeds the incoherence threshold significantly. This is because a change in phase offset will not result in a change of the intensity spectrum.

Let the operators $\mathcal{S}\{\}$ and $\int_c\{\}$ denote the transformation from the time domain to the *single sided intensity spectrum* and the *continuous Fourier spectrum* respectively (see section 7.1), and let $f^*(t)$ denote the complex conjugate of the function $f(t)$, then spectral calculus far above incoherence threshold yields for an arbitrary random function $\theta(t)$:

$$\begin{aligned}
 \mathcal{S}\{\cos(\theta(t))\} &= 2 \int_c \left\{ \langle \cos(\theta(t)) \cdot \cos^*(\theta(t+\tau)) \rangle_\infty \right\} \\
 &= 2 \int_c \left\{ \langle \frac{1}{2} \cdot \cos(\theta(t) - \theta(t+\tau)) + \frac{1}{2} \cdot \cos(\theta(t) + \theta(t+\tau)) \rangle_\infty \right\} \\
 &= \int_c \left\{ \langle \cos(\theta(t) - \theta(t+\tau)) \rangle_\infty \right\} + \int_c \left\{ \langle \cos(\theta(t) + \theta(t+\tau)) \rangle_\infty \right\} \\
 &= \int_c \left\{ \langle \cos(\theta(t) - \theta(t+\tau)) \rangle_\infty \right\} \\
 &= \int_c \left\{ \langle \exp(j \cdot \theta(t) - j \cdot \theta(t+\tau)) - j \cdot \sin(\theta(t) - \theta(t+\tau)) \rangle_\infty \right\} \\
 &= \int_c \left\{ \langle \exp(j \cdot \theta(t) - j \cdot \theta(t+\tau)) \rangle_\infty \right\} \\
 &= \int_c \left\{ \langle \exp(j \cdot \theta(t)) \cdot \exp^*(j \cdot \theta(t+\tau)) \rangle_\infty \right\}
 \end{aligned}$$

$$\boxed{\mathcal{S}\{\cos(\theta(t))\} = \frac{1}{2} \cdot \mathcal{S}\{\exp(j \cdot \theta(t))\}}$$

This expression demonstrates that the intensity spectrum of a *real* trigonometric modulation equals the intensity spectrum of a *complex* exponential function. The exponential form has no physical meaning however simplifies spectral calculation of frequency or phase modulated signals.

The intensity spectrum of a trigonometric modulated signal with offset is calculated as follows:

$$\begin{aligned}
 \mathcal{S}\{\cos(\theta(t) + \delta_0)\} &= \frac{1}{2} \cdot \mathcal{S}\{\exp(j \cdot \theta(t) + j \cdot \delta_0)\} \\
 &= \frac{1}{2} \cdot \mathcal{S}\{\exp(j \cdot \theta(t))\} \cdot |\exp(j \cdot \delta_0)|^2 \\
 &= \frac{1}{2} \cdot \mathcal{S}\{\exp(j \cdot \theta(t))\}
 \end{aligned}$$

$$\boxed{\mathcal{S}\{\cos(\theta(t) + \delta_0)\} = \mathcal{S}\{\cos(\theta(t))\}}$$

This expression demonstrates that the intensity spectrum of FM-modulated signals is invariant for constant phase shifts, when it is random modulated far above incoherence threshold. Note that this will not hold for variable phase shift.

Intensity spectrum of synthetic noise

The intensity spectrum of the synthetic noise signal is affected by random as well as periodic modulation signals $J(t)$ and $I(t)$. The addition of (periodic) modulation provides an additional phase shift, to increase the spectral width. We define two spectra: $S_h(f)$ and $\tilde{Q}(n)$.

- The continuous intensity spectrum $S_h(f)$ is associated with the random differential phase $\Delta\theta$ that is caused by the random modulation signal $J(t)$. We will refer to it as the (self) homodyne spectrum.
- The discrete Fourier spectrum $\tilde{Q}(n)$ is associated with the periodic differential phase $\Delta\phi$ that is caused by the periodic modulation signal $I(t)$ with period T . We will refer to it as the frame of the synthetic noise spectrum

These spectra are defined as follows:

$$\begin{aligned} S_h(f) &\stackrel{\text{def}}{=} \mathcal{S}\{I_{n0} \cdot \cos(\Delta\theta(t, \tau))\} &&= \text{(continuous) intensity base spectrum} \\ \tilde{Q}(n) &\stackrel{\text{def}}{=} \mathbf{j}_d^* \{\exp(j \cdot \Delta\phi(t, \tau); T)\} &&= \text{(discrete) Fourier line spectrum} \end{aligned}$$

The (self) homodyne spectrum $S_h(f)$ is the intensity spectrum of the synthetic noise in the special case that the periodic modulation is switched off.

The frame $\tilde{Q}(n)$ is a set of Fourier coefficients that is associated with $\Delta\phi$. It has no direct physical meaning, however we will use it to split a periodic FM signal in its harmonic components. The Fourier decomposition of $\exp(j \cdot \Delta\phi)$ with period $T=1/f_m=1/(2\pi\omega_n)$ has the following standard form:

$$\exp(j \cdot \Delta\phi(t, \tau)) = \sum_{n=-\infty}^{+\infty} \tilde{Q}_n \cdot \exp(j \cdot n\omega_n t)$$

Using this decomposition, and using the property that trigonometric and exponential modulated random signals yield equivalent intensity spectra, we obtain for the overall synthetic noise spectrum:

$$\begin{aligned} \mathcal{S}\{i_n(t)\} &= \mathcal{S}\{I_{n0} \cdot \cos(\Delta\alpha)\} \\ &= \mathcal{S}\{I_{n0} \cdot \cos(\alpha_\tau + \Delta\theta(t, \tau) + \Delta\phi(t, \tau))\} \\ &= \frac{1}{2} \cdot (I_{n0})^2 \cdot \mathcal{S}\{\exp(j \cdot \alpha_\tau + j \cdot \Delta\theta(t, \tau) + j \cdot \Delta\phi(t, \tau))\} \\ &= \frac{1}{2} \cdot (I_{n0})^2 \cdot \mathcal{S}\{\exp(j \cdot \alpha_\tau) \cdot \exp(j \cdot \Delta\theta(t, \tau)) \cdot \exp(j \cdot \Delta\phi(t, \tau))\} \\ &= \frac{1}{2} \cdot (I_{n0})^2 \cdot |\exp(j \cdot \alpha_\tau)|^2 \cdot \mathcal{S}\{\exp(j \cdot \Delta\theta(t, \tau)) \cdot \exp(j \cdot \Delta\phi(t, \tau))\} \\ &= \frac{1}{2} \cdot (I_{n0})^2 \cdot 1 \cdot \mathcal{S}\{\exp(j \cdot \Delta\theta(t, \tau)) \cdot \sum \tilde{Q}_n \cdot \exp(j \cdot n\omega_n t)\} \\ &= \frac{1}{2} \cdot (I_{n0})^2 \cdot \sum |\tilde{Q}_n|^2 \cdot \mathcal{S}\{\exp(j \cdot \Delta\theta(t, \tau)) \cdot \exp(j \cdot n\omega_n t)\} \end{aligned}$$

$$\mathcal{S}\{i_n(t)\} = \sum_{n=-\infty}^{+\infty} |\tilde{Q}_n|^2 \cdot S_h(f+n \cdot f_m)$$

These expressions demonstrate that the overall intensity spectrum is the superposition of similar homodyne spectra, centered at equidistant center frequencies. The frame spans the overall synthetic noise spectrum.

All comb lines of the overall synthetic noise spectrum have equal envelope, and this envelope is identical to the shape of this homodyne spectrum. As a result, the frame has

'duplicated' this homodyne spectrum to span the overall synthetic noise spectrum. The contribution to the total spectrum of all these 'duplicated' homodyne spectrum is weighted by the individual frame coefficients $\tilde{Q}_n$.

Simulation of the intensity spectra

The previous expressions are important for simulation purposes. The simulated of the line spectra (frame) were previously shown in the figures 9.5 9.7, 9.8 and 9.9. These simulations were performed in MatLab, following three succeeding steps:

- 1): $I(t)$ = array with samples of the periodical modulation current
- 2): $\Delta\phi(t, \tau) = 2\pi \cdot M \cdot \int_{-\infty}^t \{I(t) - I(t+\tau)\} \cdot dt$
- 3): $\tilde{Q}(n) = \int_a^* \{\exp(j \cdot \Delta\phi(t, \tau)); T\}$

In the first step, the current $I(t)$ was modeled by an array with samples of a single modulation period, for instance a sinusoidal or a triangular function.

In the second step, the differential phase $\Delta\phi$ was evaluated by subtracting two arrays with samples of $I(t)$ and $I(t+\tau)$, and then elementwise summated to simulate the integration.

In the third step, the spectral components of $\tilde{Q}(n)$ were extracted from the samples of $\Delta\phi$ using standard FFT techniques. This approach is applicable for arbitrary modulation currents, and therefore more flexible than an approach with Bessel functions. Moreover, FFT techniques produce the same result with less computational effort.

The homodyne spectrum was made available by measurement (figure 9.15) or by calculation with many random numbers (figure 9.16). The simulation was performed, similarly as described above. The intensity spectrum was evaluated by averaging the square of many FFT samples in a small frequency band.

- 1): $J(t)$ = array with Gaussian distributed random numbers
- 2): $\Delta\theta(t, \tau) = 2\pi \cdot M \cdot \int_{-\infty}^t \{J(t) - J(t+\tau)\} \cdot dt$
- 3): $S_h(f) = \mathcal{F}\{I_{n0} \cdot \cos(\Delta\theta(t, \tau))\}$

The major difference is that random numbers were used in stead of a periodic signal, which requires many samples to obtain a smoothed intensity spectrum. A single FFT calculation over $15 \cdot 10^6$ numbers was avoided using 450 individual FFT calculations over 32768 samples. All these intermediate results were combined in a common average.

Power conservation

Using the well-known Parseval identity for spectra, we obtain for the overall homodyne power and overall frame power:

$$\begin{aligned} \int_0^{+\infty} |S_h(f)|^2 \cdot df &= \left\langle |I_{n0} \cdot \cos(\Delta\theta(t, \tau))|^2 \right\rangle_{\infty} = \frac{1}{2} \cdot (I_{n0})^2 = (i_{\text{rms}})^2 \\ \sum_{n=-\infty}^{+\infty} \tilde{Q}_n &= \left\langle |\exp(j \cdot \Delta\phi(t, \tau))|^2 \right\rangle_{\infty} = 1 \end{aligned}$$

$$\begin{aligned} \int_0^{+\infty} |S_h(f)|^2 \cdot df &= (i_{\text{rms}})^2 \\ \sum_{n=-\infty}^{+\infty} \tilde{Q}_n &= 1 \end{aligned}$$

These expressions demonstrate that the frame spreads the homodyne spectrum out over a wide frequency interval, however preserves all available power. The wider the envelope of the frame, the lower the overall spectral density will be.

9.3.4. Statistical analysis far above incoherence threshold

The statistical distribution of synthetic noise differs significantly from the distribution of natural noise. The maximum momentaneous magnitude of pure synthetic noise will never exceed twice the rms-value, while the spikes in natural noise will exceed this value with ease.

Above incoherence threshold the distribution of synthetic noise becomes insensitive for the distribution of the random modulation. This subsection demonstrates that this distribution approaches to a trigonometric distribution, far above the incoherence threshold.

The output signal of a synthetic noise generator has the following general form:

$$\begin{aligned} I_n(t) &= I_{n0} \cdot (\cos(\Delta\alpha(t, \tau)) + 1) \\ i_n(t) &= I_{n0} \cdot \cos(\Delta\alpha(t, \tau)) \end{aligned}$$

The differential phase $\Delta\alpha$ is not restricted in magnitude, and may have any statistical distribution and momentaneous magnitude. The wrapped phase $\varepsilon \triangleq \arccos(\cos(\Delta\alpha))$ is a quantity that is restricted from 0 to π . Modulation with $\Delta\alpha$ or ε will result in equal synthetic noise signals however the distribution function of both phase terms will differ significantly.

The higher the incoherence is, the higher the rms-value of $\Delta\alpha$ will be, and the more ε is wrapped. In the limit, all values between $(0 \leq \varepsilon \leq \pi)$ have equal statistical performance, which gives ε a uniform distribution. As a result, the statistical distribution of the wrapped phase is $p_{\varepsilon}(\varepsilon) = 1/\pi$ within that interval, and is zero for $(\varepsilon < 0)$ and $(\varepsilon > \pi)$.

Let $i_n(t)$ be a synthetic noise current, and let i_{nx} be an arbitrary constant value within the interval $(-I_{n0} \leq i_n \leq +I_{n0})$, then $\varepsilon(t)$ and ε_x are the associated wrapped phase terms. The probability that $i_n(t)$ exceeds the value i_{nx} is then as follows:

$$\begin{aligned}
\underline{\underline{R}}\{i_n > i_{nx}\} &= \underline{\underline{R}}\{\arccos(i_n/I_{n0}) < \arccos(i_{nx}/I_{n0})\} \\
&= \underline{\underline{R}}\{\varepsilon < \varepsilon_x\} \\
&= 1 - \underline{\underline{R}}\{\varepsilon > \varepsilon_x\} \\
&= 1 - (\varepsilon_x/\pi) \\
&= 1 - \arccos(i_{nx}/I_{n0})/\pi
\end{aligned}$$

The distribution function, or probability density, is the derivative of this probability, and is as follows:

$$p_i(i_{nx}) = \frac{\partial}{\partial i_{nx}} \{1 - \arccos(i_{nx}/I_{n0})/\pi\} = \frac{1/\pi}{\sqrt{I_{n0}^2 - i_{nx}^2}}$$

$p_i(i) = \frac{1/\pi}{\sqrt{2 \cdot i_{rms}^2 - i^2}} \quad i \leq i_{rms} \cdot \sqrt{2}$

This expression demonstrates that synthetic noise has a trigonometric distribution, as was shown in figure 9.19. It differs significantly from Gaussian distributed noise, which has the following well-known form:

$$p_i(i) = \frac{1}{\sqrt{2 \cdot \pi} \cdot i_{rms}} \cdot \exp(-1/2 \cdot (i/i_{rms})^2) \quad \text{Gaussian (normal) distribution}$$

The distribution function is a commonly used alternative for spectral calculus to evaluate the mean value or rms-value of noise. It has the following general properties:

$$\begin{aligned}
\underline{\underline{R}}\{I > I_x\} &= \int_{-\infty}^{+I_x} p(i) \cdot di && \text{probability} \\
\langle I(t) \rangle_{\infty} &= \int_{-\infty}^{+\infty} p(i) \cdot i \cdot di && \text{mean value} \\
\langle I(t)^2 \rangle_{\infty} &= \int_{-\infty}^{+\infty} p(i) \cdot i^2 \cdot di && \text{intensity}
\end{aligned}$$

9.3.5. Conclusions

In conclusion, we developed a general theory for the generation of synthetic noise. We proposed an adequate calculation model and developed a time-domain description for the synthetic noise current. These formulae demonstrate that the magnitude of a synthetic noise current spans a restricted range.

Our theoretical study has resulted in the following concepts and expressions:

- Proposal of a new concept, the *incoherence threshold*. A well-designed synthetic noise generator must always operate far above this threshold.

Our definition of operation above coherence threshold is equivalent with the condition that the differential delay time of the interferometer must exceed the coherence time of the laser [907]. A shortage in differential delay is fully compensatable with additional noise injection. This addition facilitate operation above incoherence threshold.

- Derivation of expressions for the rms-magnitude, the intensity spectrum S_i and the noise-current ratio μ of synthetic noise currents. The validation of these expressions is restricted to synthetic noise generators operating far above incoherence threshold.
- Derivation of expressions for the incoherence, in case of pink noise and white noise injection. These expressions resulted in conditions to ensure operation above incoherence threshold.
- Derivation of an expression for the overall synthetic noise spectrum. This spectrum is the superposition of many similar homodyne spectra, centered at equidistant center frequencies (comb lines). The frame spans the overall spectrum, and weights the contribution of each comb line to the overall spectrum. These expressions facilitate simulation of the synthetic noise spectrum, for arbitrary modulation currents.
- Evaluated of the probability distribution function of synthetic noise. It is demonstrated that this statistical distribution differs significantly from Gaussian distributed noise.

This theoretical section 9.3 is essential for an analysis as described in section 9.2.

Appendix

A. Transformation rules for two-port parameters

Two-port parameters (see section 2.2) are usually measured in s-parameters, normalized to a *real* reference impedance $Z_0=R_0=1/G_0$. The most relevant parameter transformations are therefore transformations from or to the s-parameter format. They are provided below.

For two-port matrix parameters (Z,Y,A,H,T,S) see section 2.3.1. For two-port virtual circuit parameters (BJT,FET) see section 2.3.2.

	<i>from S</i>	<i>to S (via Y)</i>
Z	$q_{sz} = ((s_{11}-1) \cdot (s_{22}-1) - s_{12} \cdot s_{21}) / (2 \cdot R_0)$ $z_{11} = (1 - s_{22}) / q_{sz} - R_0$ $z_{22} = (1 - s_{11}) / q_{sz} - R_0$ $z_{21} = s_{21} / q_{sz}$ $z_{12} = s_{12} / q_{sz}$	$q_{zs} = ((z_{11} + R_0) \cdot (z_{22} + R_0) - z_{12} \cdot z_{21}) / (2 \cdot R_0)$ $s_{11} = -(R_0 + z_{22}) / q_{zs} + 1$ $s_{22} = -(R_0 + z_{11}) / q_{zs} + 1$ $s_{21} = z_{21} / q_{zs}$ $s_{12} = z_{12} / q_{zs}$
Y	$q_{sy} = ((s_{11}+1) \cdot (s_{22}+1) - s_{21} \cdot s_{12}) \cdot (R_0/2)$ $y_{11} = (1 + s_{22}) / q_{sy} - 1/R_0$ $y_{22} = (1 + s_{11}) / q_{sy} - 1/R_0$ $y_{21} = -s_{21} / q_{sy}$ $y_{12} = -s_{12} / q_{sy}$	$q_{ys} = ((y_{11} + G_0) \cdot (y_{22} + G_0) - y_{21} \cdot y_{12}) \cdot (R_0/2)$ $s_{11} = (G_0 + y_{22}) / q_{ys} - 1$ $s_{22} = (G_0 + y_{11}) / q_{ys} - 1$ $s_{21} = -y_{21} / q_{ys}$ $s_{12} = -y_{12} / q_{ys}$
A	$q_{sa} = ((s_{11}+1) \cdot (s_{22}+1) - s_{12} \cdot s_{21}) / 2$ $a_{11} = (1 + s_{11} - q_{sa}) / s_{21}$ $a_{22} = (1 + s_{22} - q_{sa}) / s_{21}$ $a_{21} = (-s_{11} - s_{22} + q_{sa}) / s_{21} / R_0$ $a_{12} = (q_{sa} / s_{21} \cdot R_0)$	$q_{as} = (a_{11} + a_{12} / R_0 + a_{21} \cdot R_0 + a_{22}) / 2$ $s_{11} = (a_{11} + a_{12} / R_0) / q_{as} - 1$ $s_{22} = (a_{22} + a_{12} / R_0) / q_{as} - 1$ $s_{21} = 1 / q_{as}$ $s_{12} = (a_{11} \cdot a_{22} - a_{21} \cdot a_{12}) / q_{as}$
H	$q_{sh} = ((s_{11}-1) \cdot (s_{22}+1) - s_{12} \cdot s_{21}) / 2$ $h_{11} = -((1 + s_{22}) / q_{sh} + 1) \cdot R_0$ $h_{22} = -((1 - s_{11}) / q_{sh} + 1) / R_0$ $h_{21} = +s_{21} / q_{sh}$ $h_{12} = -s_{12} / q_{sh}$	$q_{hs} = ((s_{11} / R_0 + 1) \cdot (s_{22} \cdot R_0 + 1) - s_{12} \cdot s_{21}) / 2$ $s_{11} = -(1 + h_{22} \cdot R_0) / q_{hs} + 1$ $s_{22} = +(1 + h_{11} / R_0) / q_{hs} - 1$ $s_{21} = -h_{21} / q_{hs}$ $s_{12} = +h_{12} / q_{hs}$
T	$\Delta_s = (s_{11} \cdot s_{22} - s_{21} \cdot s_{12})$ $t_{11} = 1 / s_{21}$ $t_{22} = -\Delta_s / s_{21}$ $t_{21} = +s_{11} / s_{21}$ $t_{12} = -s_{22} / s_{21}$	$\Delta_t = (t_{11} \cdot t_{22} - t_{21} \cdot t_{12})$ $s_{11} = t_{21} / t_{11}$ $s_{22} = -t_{12} / t_{11}$ $s_{21} = 1 / t_{11}$ $s_{12} = \Delta_t / t_{11}$
B J T	$q_{sy} = ((s_{11}+1) \cdot (s_{22}+1) - s_{21} \cdot s_{12}) \cdot (R_0/2)$ $\alpha_e = (s_{12} - s_{21}) / (1 + s_{22} - s_{12} - q_{sy} / R_0)$ $z_e = q_{sy} / (1 + s_{22} - s_{21} - q_{sy} / R_0)$ $z_{ce} = q_{sy} / (1 + s_{11} - s_{12} - q_{sy} / R_0)$ $z_{cb} = q_{sy} / (s_{12})$	$y_{11} = 1 / (1 + \alpha_e) / z_e + 1 / z_{cb}$ $y_{22} = 1 / z_{ce} + 1 / z_{cb}$ $y_{21} = \alpha_e / (1 + \alpha_e) / z_e - 1 / z_{cb}$ $y_{12} = -1 / z_{cb}$
F E T	$q_{sy} = ((s_{11}+1) \cdot (s_{22}+1) - s_{21} \cdot s_{12}) \cdot (R_0/2)$ $g_m = (s_{12} - s_{21}) / q_{sy}$ $z_{gs} = q_{sy} / (1 + s_{22} - s_{12} - q_{sy} / R_0)$ $z_{ds} = q_{sy} / (1 + s_{11} - s_{12} - q_{sy} / R_0)$ $z_{dg} = q_{sy} / (s_{12})$	$y_{11} = 1 / z_{gs} + 1 / z_{dg}$ $y_{22} = 1 / z_{ds} + 1 / z_{dg}$ $y_{21} = g_m - 1 / z_{dg}$ $y_{12} = -1 / z_{dg}$

$$G_0=1/R_0=1/Z_0$$

Appendix

B. Algorithms for overdetermined matrix divisions

Matrix division is a generalization of matrix inversion, and is used in many chapters of this manuscript. It is defined for square as well as rectangular matrices, provided that the associated matrix dimensions match. In all probability, the definition originates from the MATLAB[®][123] manuals, although it should be defined in any standard textbook discussion on linear algebra. To simplify matters,

$$\mathbf{A} \backslash \mathbf{B} \approx (\mathbf{A}^{-1}) \cdot \mathbf{B} \quad \text{left-hand matrix division}$$

$$\mathbf{A} / \mathbf{B} \approx \mathbf{A} \cdot (\mathbf{B}^{-1}) \quad \text{right-hand matrix division}$$

The advantage of using division instead of inversion is that the concept is more general. It simplifies mathematical expressions when solving *overdetermined* sets of linear equations. This appendix describes the mathematical background of matrix division.

B.1. Definition of general matrix division.

The general definition is somewhat complicated, and is not representative for the way the division is evaluated (see section B.5) The division between two matrices is defined as:

$$\begin{array}{l} \mathbf{A} \backslash \mathbf{b} \stackrel{\text{def}}{=} (\mathbf{A}' \cdot \mathbf{A})^{-1} \cdot (\mathbf{A}' \cdot \mathbf{b}) \\ \mathbf{c} / \mathbf{A} \stackrel{\text{def}}{=} (\mathbf{c} \cdot \mathbf{A}') \cdot (\mathbf{A} \cdot \mathbf{A}')^{-1} \end{array} \Rightarrow \mathbf{A} \backslash \mathbf{b} = (\mathbf{b}' / \mathbf{A}')'$$

$\mathbf{A}' \equiv \text{conj}(\mathbf{A}^T) = \text{conjugated transpose}$

The definition holds for square as well for rectangular matrices, provided that the associated matrix dimensions match. Its validation includes complex matrices. The next subsections will illustrate the meaning of these definitions.

B.2. Exact division of a square matrix

A classic application of matrix division and matrix inversion is finding a solution of a set of linear equations. These equation sets, have $\mathbf{A} \cdot \mathbf{z} = \mathbf{b}$ as general form. In this equation represent matrix $\mathbf{A}$ and $\mathbf{b}$ various equation constants and is $\mathbf{z}$ the solution vector. Matrix $\mathbf{A}$ must be square and invertable to enable $\mathbf{z} = \mathbf{A}^{-1} \cdot \mathbf{b}$.

In this special case the general definitions of matrix division simplifies significantly since $(\mathbf{A}' \cdot \mathbf{A})^{-1} \cdot \mathbf{A}' \equiv \mathbf{A}^{-1}$. This simplification emphasizes that matrix division is a generalization of matrix inversion because:

$$\mathbf{A} \backslash \mathbf{b} = (\mathbf{A}^{-1}) \cdot \mathbf{b}$$

$$\mathbf{c} / \mathbf{A} = \mathbf{c} \cdot (\mathbf{A}^{-1})$$

$$\mathbf{c} \cdot (\mathbf{A} \backslash \mathbf{b}) = \mathbf{c} \cdot (\mathbf{A}^{-1}) \cdot \mathbf{b} = (\mathbf{c} / \mathbf{A}) \cdot \mathbf{b}$$

The vector $\mathbf{z}_1 = \mathbf{A} \backslash \mathbf{b}$ is the solution of matrix equation $\mathbf{A} \cdot \mathbf{z}_1 = \mathbf{b}$.
 The vector $\mathbf{z}_2 = \mathbf{b} / \mathbf{A}$ is the solution of matrix equation $\mathbf{z}_2 \cdot \mathbf{A} = \mathbf{b}$.

Many textbook discussions on linear algebra suggest that the evaluation of $\mathbf{A} \backslash \mathbf{B}$ requires the inversion of matrix $\mathbf{A}$ followed by a matrix product with $\mathbf{B}$. This is as inefficient as calculating: $8/4 = 8 \cdot (1/4) = 8 \cdot (0.25) = 2$. Software packages, such as Matlab or LinPack

are provided with routines to calculate $\mathbf{A} \backslash \mathbf{b}$ more efficient than calculating $(\mathbf{A}^{-1}) \cdot \mathbf{b}$ (see section B.5).

B.3. Least squares overdetermined division of a rectangular matrix

An important application of the generalized matrix division is finding a fair solution for overdetermined sets of equations. An overdetermined equation set, has $\mathbf{A} \cdot \mathbf{z} \approx \mathbf{b}$ as general form, and has no exact 'solution'. In this 'equation' represent matrix $\mathbf{A}$ and $\mathbf{b}$ various equation constants and $\mathbf{z}$ one of the many 'solution' vectors. Matrix $\mathbf{A}$ is a rectangular matrix.

The usefulness of 'solutions' $\mathbf{z}$ is closely related with the deviation $\Delta \mathbf{b}$ of vector $\mathbf{A} \cdot \mathbf{z}$ from vector $\mathbf{b}$. This deviation can be analyzed as follows:

$$\mathbf{A} \cdot \mathbf{z} = \mathbf{b} + \Delta \mathbf{b}$$

$$\text{let } \Delta \mathbf{z} \stackrel{\text{def}}{=} \mathbf{A} \backslash \mathbf{b} - \mathbf{z} \quad = \text{an offset vector of solution } \mathbf{z} \text{ relative to vector } \mathbf{A} \backslash \mathbf{b}$$

$$\text{let } \boldsymbol{\beta} \stackrel{\text{def}}{=} \mathbf{A} \cdot (\mathbf{A} \backslash \mathbf{b}) - \mathbf{b} \quad = \text{a substitution vector to simplify the equations}$$

$$\Delta \mathbf{b} = \mathbf{A} \cdot \mathbf{z} - \mathbf{b}$$

$$\Delta \mathbf{b} = \mathbf{A} \cdot (\mathbf{A} \backslash \mathbf{b} - \Delta \mathbf{z}) - \mathbf{b}$$

$$\Delta \mathbf{b} = \mathbf{A} \cdot (\mathbf{A} \backslash \mathbf{b}) - \mathbf{A} \cdot (\Delta \mathbf{z}) - \mathbf{b}$$

$$\Delta \mathbf{b} = \boldsymbol{\beta} - \mathbf{A} \cdot \Delta \mathbf{z}$$

The deviation vector $\Delta \mathbf{b}$ is a function of the offset vector $\Delta \mathbf{z}$. When using the relation $\mathbf{A}' \cdot \boldsymbol{\beta} = \mathbf{0}$, which can be demonstrated simply from the definitions of $\boldsymbol{\beta}$ and $\mathbf{A} \backslash \mathbf{b}$, the value $|\Delta \mathbf{b}|^2$ equals:

$$|\Delta \mathbf{b}|^2 = (\boldsymbol{\beta} - \mathbf{A} \cdot \Delta \mathbf{z})' \cdot (\boldsymbol{\beta} - \mathbf{A} \cdot \Delta \mathbf{z})$$

$$|\Delta \mathbf{b}|^2 = \boldsymbol{\beta}' \cdot \boldsymbol{\beta} + (\mathbf{A} \cdot \Delta \mathbf{z})' \cdot (\mathbf{A} \cdot \Delta \mathbf{z}) - \boldsymbol{\beta}' \cdot (\mathbf{A} \cdot \Delta \mathbf{z}) - (\mathbf{A} \cdot \Delta \mathbf{z})' \cdot \boldsymbol{\beta}$$

$$|\Delta \mathbf{b}|^2 = |\boldsymbol{\beta}|^2 + |\mathbf{A} \cdot \Delta \mathbf{z}|^2 - (\boldsymbol{\beta}' \cdot \mathbf{A} \cdot \Delta \mathbf{z}) - (\Delta \mathbf{z}' \cdot \mathbf{A}' \cdot \boldsymbol{\beta})$$

$$|\Delta \mathbf{b}|^2 = |\boldsymbol{\beta}|^2 + |\mathbf{A} \cdot \Delta \mathbf{z}|^2 - 0 - 0$$

This analysis shows that in all situations $|\Delta \mathbf{b}|^2 \geq |\boldsymbol{\beta}|^2$, and that $|\Delta \mathbf{b}|^2$ is minimal when the offset vector $\Delta \mathbf{z} = \mathbf{0}$. This property makes $\mathbf{z} = \mathbf{A} \backslash \mathbf{b}$ often a fair choice, since it minimizes the 'error' in a least squares sense.

The solution $\mathbf{z}_1 = \mathbf{A} \backslash \mathbf{b}$ of the matrix equation $\mathbf{A} \cdot \mathbf{z}_1 = \mathbf{b} + \Delta \mathbf{b}_1$ minimizes $|\Delta \mathbf{b}_1|^2$
 The solution $\mathbf{z}_2 = \mathbf{b} / \mathbf{A}$ of the matrix equation $\mathbf{z}_2 \cdot \mathbf{A} = \mathbf{b} + \Delta \mathbf{b}_2$ minimizes $|\Delta \mathbf{b}_2|^2$

These properties hold for real matrices as was for complex matrix equations (with complex solutions),

B.4. Least squares overdetermined division, restricted to real results

For many physical applications, complex solutions are excluded. Over-determined linear equation sets, have $\mathbf{A} \cdot \mathbf{x} \approx \mathbf{b}$ as general form, in which the matrices $\mathbf{A}$ and $\mathbf{b}$ may be *complex* provided that $\mathbf{x}$ is real. When complex solutions are excluded, the vector $\mathbf{A} \backslash \mathbf{b}$ is usually excluded to as solution.

The usefulness of real 'solutions' $\mathbf{x}$ is closely related with the complex deviation $\Delta\mathbf{b}$ of vector $\mathbf{A}\cdot\mathbf{x}$ from vector $\mathbf{b}$. This complex deviation can be transformed into a real format as follows:

$$\begin{aligned}\mathbf{A}\cdot\mathbf{x} &= \mathbf{b} + \Delta\mathbf{b} \\ (\mathbf{A}_{re} + j\cdot\mathbf{A}_{im})\cdot\mathbf{x} &= (\mathbf{b}_{re} + j\cdot\mathbf{b}_{im}) + (\Delta\mathbf{b}_{re} + j\cdot\Delta\mathbf{b}_{im}) \\ \begin{bmatrix} \mathbf{A}_{re} \\ \mathbf{A}_{im} \end{bmatrix} \cdot \mathbf{x} &= \begin{bmatrix} \mathbf{b}_{re} \\ \mathbf{b}_{im} \end{bmatrix} + \begin{bmatrix} \Delta\mathbf{b}_{re} \\ \Delta\mathbf{b}_{im} \end{bmatrix} \\ \Rightarrow \left| \frac{\Delta\mathbf{b}_{re}}{\Delta\mathbf{b}_{im}} \right|^2 &= |\Delta\mathbf{b}_{re}|^2 + |\Delta\mathbf{b}_{im}|^2 = |\Delta\mathbf{b}_{re} + j\cdot\Delta\mathbf{b}_{im}|^2 = |\Delta\mathbf{b}|^2\end{aligned}$$

This transformed form illustrates that vector $\Delta\mathbf{b}$ and the transformed variant of this vector have equal magnitude. This means that a real solution vector $\mathbf{x}$ that minimizes the transformed equation 'error' also minimizes the complex equation (within the restriction of real solutions).

The <i>real</i> solution vector $\mathbf{x}$ that minimizes $ \Delta\mathbf{b} ^2$ of the complex matrix equation $\mathbf{A}\cdot\mathbf{x}_1=\mathbf{b}+\Delta\mathbf{b}$, equals:	$\mathbf{x}_1 = \begin{bmatrix} \mathbf{A}_{re} \\ \mathbf{A}_{im} \end{bmatrix} \setminus \begin{bmatrix} \mathbf{b}_{re} \\ \mathbf{b}_{im} \end{bmatrix}$
The <i>real</i> solution vector $\mathbf{x}$ that minimizes $ \Delta\mathbf{c} ^2$ of the complex matrix equation $\mathbf{x}_2\cdot\mathbf{A}=\mathbf{c}+\Delta\mathbf{c}$, equals:	$\mathbf{x}_2 = \begin{bmatrix} \mathbf{c}_{re} \\ \mathbf{c}_{im} \end{bmatrix} / \begin{bmatrix} \mathbf{A}_{re} \\ \mathbf{A}_{im} \end{bmatrix}$

B.5. Division algorithm

Although the definition of the matrix division has facilitated a direct evaluation, a more appropriate approach is recommended. In the general case that $\mathbf{A}$ is rectangular, the definition is reduced by means of the QR-decomposition [413,414] of matrix $\mathbf{A}$ in an orthogonal matrix $\mathbf{Q}$ and an upper triangle matrix $\mathbf{R}$.

Because $\mathbf{A} = \mathbf{Q}\cdot\mathbf{R}$ and $\mathbf{Q}'\cdot\mathbf{Q} = \mathbf{1}$ by definition, the evaluation of the matrix division reduces to:

$$\begin{aligned}\mathbf{A}\backslash\mathbf{b} &= (\mathbf{R}'\cdot\mathbf{Q}' \cdot \mathbf{Q}\cdot\mathbf{R})^{-1} \cdot (\mathbf{R}'\cdot\mathbf{Q}'\cdot\mathbf{b}) \\ &= (\mathbf{R}'\cdot\mathbf{R})^{-1} \cdot (\mathbf{R}'\cdot\mathbf{Q}') \cdot \mathbf{b} \\ &= \mathbf{R} \setminus (\mathbf{Q}'\cdot\mathbf{b})\end{aligned}$$

The benefit of this approach is that the upper triangle form of $\mathbf{R}$ dramatically simplifies the division, because many elements are zero. If back substitution ([413:section 2.2] is performed, then an explicit division with $\mathbf{R}$ is not required. Further, a robust QR-factorization, based on Householder transformations ([413:section 11.2, 414] reduces numerical round-off errors for ill-conditioned matrices $\mathbf{A}$. And finally, the explicit evaluation of $\mathbf{Q}$ is not required if each Householder transformation step, that reduces $\mathbf{A}$ toward $\mathbf{R}$, is simultaneously performed on vector $\mathbf{b}$.

Appendix

C. Algorithm for polynomial curve fits

This appendix discusses an algorithm for finding the *real* coefficients of a polynomial that optimally fits to a table with measured data. The polynomial represents a transfer function of the following general form:

$$H(j\omega) = H(s) \stackrel{\text{def}}{=} a_0 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_n \cdot s^n$$

When an *exact* fit exists it is irrelevant how the coefficients $[a_0, a_1, a_2, \dots, a_n]$ are calculated, since that solution is unique. When $H(s)$ is measured at k frequencies ($k > n$) numerous 'solutions' exist. We call a solution *optimal* when the *relative* error δ of $H(s)$ is minimized, rms-averaged over all frequencies of interest.

Let h denote the measured function value of $H(s)$ at complex frequency $s=j\omega$, then:

$$a_0 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_n \cdot s^n \approx h$$

$$a_0 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_n \cdot s^n = h \cdot (1 + \delta)$$

$$a_0/h + a_1 \cdot s/h + a_2 \cdot s^2/h + \dots + a_n \cdot s^n/h = (1 + \delta)$$

$$a_0 \cdot (w/h) + a_1 \cdot s \cdot (w/h) + a_2 \cdot s^2 \cdot (w/h) + \dots + a_n \cdot s^n \cdot (w/h) = w \cdot (1 + \delta)$$

The quantity w represents an arbitrary weighing factor, to be discussed below. To simplify these equations, the vectors α , R , ΔR , and Q are as follows defined for all frequencies in the table:

$$\alpha \stackrel{\text{def}}{=} \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_n \end{bmatrix} \quad R \stackrel{\text{def}}{=} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \dots \\ w_k \end{bmatrix} \quad \Delta R \stackrel{\text{def}}{=} \begin{bmatrix} w_1 \cdot \delta_1 \\ w_2 \cdot \delta_2 \\ w_3 \cdot \delta_3 \\ \dots \\ w_k \cdot \delta_k \end{bmatrix}$$

$$Q \stackrel{\text{def}}{=} \begin{bmatrix} w_1/h_1 & w_1/h_1 \cdot s_1 & w_1/h_1 \cdot s_1^2 & \dots & w_1/h_1 \cdot s_1^n \\ w_2/h_2 & w_2/h_2 \cdot s_2 & w_2/h_2 \cdot s_2^2 & \dots & w_2/h_2 \cdot s_2^n \\ w_3/h_3 & w_3/h_3 \cdot s_3 & w_3/h_3 \cdot s_3^2 & \dots & w_3/h_3 \cdot s_3^n \\ \dots & \dots & \dots & \dots & \dots \\ w_k/h_k & w_k/h_k \cdot s_k & w_k/h_k \cdot s_k^2 & \dots & w_k/h_k \cdot s_k^n \end{bmatrix}$$

$$Q \cdot \alpha = R + \Delta R$$

$$\alpha = \begin{bmatrix} Q_{\text{re}} \\ Q_{\text{im}} \end{bmatrix} \setminus \begin{bmatrix} R_{\text{re}} \\ R_{\text{im}} \end{bmatrix}$$

see appendix **B** for the definition of an over-determined left-hand matrix division, restricted to real results

The vector α is a solution that minimizes ΔR in a least squares sense (see appendix **B**). This means that the rms-average of all weighed 'errors' ($w \cdot \delta$) is minimized.

When the weighing factors w are *balanced*, meaning that they are all equal for all frequencies (e.g. $w=1$), the relative error δ is minimized and α is the requested optimal solution. The use of *unbalanced* weighing factors w makes the fit more (or less) sensitive in restricted frequency ranges. In case $w=0$ in a frequency interval, the associated errors are completely ignored. This is handy to exclude spikes and other errors in measurements, or to exclude frequency ranges in which the transfer function is inadequately predicted by the chosen polynomial.

Appendix

D. Algorithm for rational curve fits

This appendix discusses an algorithm for finding the *real* coefficients of a rational function that optimally fits to a table with measured data. The rational function represents a transfer function of the following general form:

$$H(j\omega) = H(s) \stackrel{\text{def}}{=} g \cdot \frac{a_0 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{m-1} \cdot s^{m-1} + s^m}{b_0 + b_1 \cdot s + b_2 \cdot s^2 + \dots + b_{n-1} \cdot s^{n-1} + s^n} = g \cdot \frac{T(s)}{N(s)}$$

When an *exact* fit exists it is irrelevant how the coefficients are calculated, since that solution is unique. When $H(s)$ is measured at k frequencies ($k > n$) numerous 'solutions' exist. We call a solution *optimal* when the *relative* error δ of $H(s)$ is minimized, rms-averaged over all frequencies of interest.

Let h denote the measured function value of $H(s)$ at complex frequency $s=j\omega$, and w an arbitrary weighing factor as discussed in appendix C then:

$$g \cdot \frac{a_0 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{m-1} \cdot s^{m-1} + s^m}{b_0 + b_1 \cdot s + b_2 \cdot s^2 + \dots + b_{n-1} \cdot s^{n-1} + s^n} \approx h$$

$$g \cdot \frac{a_0 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{m-1} \cdot s^{m-1} + s^m}{b_0 + b_1 \cdot s + b_2 \cdot s^2 + \dots + b_{n-1} \cdot s^{n-1} + s^n} = h \cdot (1 + \delta)$$

$$w \cdot g \cdot \frac{a_0 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{m-1} \cdot s^{m-1} + s^m}{b_0 + b_1 \cdot s + b_2 \cdot s^2 + \dots + b_{n-1} \cdot s^{n-1} + s^n} = h \cdot w \cdot (1 + \delta)$$

$$w \cdot g \cdot (a_0 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{m-1} \cdot s^{m-1} + s^m) = h \cdot w \cdot (1 + \delta) (b_0 + b_1 \cdot s + b_2 \cdot s^2 + \dots + b_{n-1} \cdot s^{n-1} + s^n)$$

$$\eta \stackrel{\text{def}}{=} N(s) \stackrel{\text{def}}{=} (b_0 + b_1 \cdot s + b_2 \cdot s^2 + \dots + b_{n-1} \cdot s^{n-1} + s^n)$$

$$w \cdot g \cdot (a_0 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{m-1} \cdot s^{m-1} + s^m) - h \cdot w \cdot (b_0 + b_1 \cdot s + b_2 \cdot s^2 + \dots + b_{n-1} \cdot s^{n-1}) = h \cdot w \cdot (s^n + \delta \cdot \eta)$$

$$w/\eta \cdot g/h \cdot (a_0 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{m-1} \cdot s^{m-1} + s^m) - w/\eta \cdot (b_0 + b_1 \cdot s + b_2 \cdot s^2 + \dots + b_{n-1} \cdot s^{n-1}) = w/\eta \cdot s^n + w \cdot \delta$$

To simplify these equations, the vectors α , R , ΔR and Q are as follows defined for all frequencies in the table:

$$\alpha \stackrel{\text{def}}{=} \begin{bmatrix} g \cdot a_0 \\ g \cdot a_1 \\ \dots \\ g \cdot a_{m-1} \\ g \\ b_0 \\ b_1 \\ \dots \\ b_{n-1} \end{bmatrix} \quad R \stackrel{\text{def}}{=} \begin{bmatrix} w_1/\eta_1 \cdot s_1^n \\ w_2/\eta_2 \cdot s_2^n \\ w_3/\eta_3 \cdot s_3^n \\ \dots \\ w_k/\eta_k \cdot s_k^n \end{bmatrix} \quad \Delta R \stackrel{\text{def}}{=} \begin{bmatrix} w_1 \cdot \delta_1 \\ w_2 \cdot \delta_2 \\ w_3 \cdot \delta_3 \\ \dots \\ w_k \cdot \delta_k \end{bmatrix} \quad W \stackrel{\text{def}}{=} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \dots \\ w_k \end{bmatrix}$$

$$Q \stackrel{\text{def}}{=} \begin{bmatrix} w_1/\eta_1/h_1 & w_1/\eta_1/h_1 \cdot s_1 & \dots & w_1/\eta_1/h_1 \cdot s_1^m & -w_1/\eta_1 & -w_1/\eta_1 \cdot s_1 & \dots & -w_1/\eta_1 \cdot s_1^{n-1} \\ w_2/\eta_2/h_2 & w_2/\eta_2/h_2 \cdot s_2 & \dots & w_2/\eta_2/h_2 \cdot s_2^m & -w_2/\eta_2 & -w_2/\eta_2 \cdot s_2 & \dots & -w_2/\eta_2 \cdot s_2^{n-1} \\ w_3/\eta_3/h_3 & w_3/\eta_3/h_3 \cdot s_3 & \dots & w_3/\eta_3/h_3 \cdot s_3^m & -w_3/\eta_3 & -w_3/\eta_3 \cdot s_3 & \dots & -w_3/\eta_3 \cdot s_3^{n-1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ w_k/\eta_k/h_k & w_k/\eta_k/h_k \cdot s_k & \dots & w_k/\eta_k/h_k \cdot s_k^m & -w_k/\eta_k & -w_k/\eta_k \cdot s_k & \dots & -w_k/\eta_k \cdot s_k^{n-1} \end{bmatrix}$$

$$Q \cdot \alpha = R + \Delta R$$

$$\alpha = \begin{bmatrix} Q_{re} \\ Q_{im} \end{bmatrix} \setminus \begin{bmatrix} R_{re} \\ R_{im} \end{bmatrix}$$

see appendix B for the definition of an over-determined left-hand matrix division, restricted to real results

The vector α is a solution that minimizes ΔR in a least squares sense (see appendix B). This means that the rms-average of all weighed 'errors' ($w \cdot \delta$) is minimized.

This approach of calculating α seems to be somewhat tricky because the calculation of the values η requires the coefficients of α as input. The essential point of this approach is that the numerical values of η are *estimated* for each frequency of the table. If an exact fit exists the solution α is independent of the estimated values η . If it does not exist, the estimated values of η are iterative improved until a stable solution vector α is found. The iteration algorithm is as symbolically illustrated below:

$\alpha = \text{TransferFit}(\mathbf{h}, \mathbf{s}, n, m, \mathbf{w})$	
$\mathbf{h}, \mathbf{s}$ are column vectors with length k representing the (measured) transfer function values and the associated complex frequencies.	
The integers n and m represent the requested polynomial orders (number of poles and zeros).	
$\mathbf{w}$ is a column vector representing the weighing factors; when omitted choose $\mathbf{w} \equiv 1$	
$s_{\text{avg}} = j \cdot \sqrt[k]{ s_1 \cdot s_2 \cdot s_3 \cdot s_4 \cdot \dots \cdot s_k }$	{ is the geometric mean of all complex frequencies $s = j\omega$ }
$\eta = \frac{1 + (s/s_{\text{avg}})^m}{h}$	{ is a fair initial guess }
$\alpha_0 := \mathbf{0}$;	
<u>repeat</u>	
evaluate $[\mathbf{R}, \mathbf{Q}]$	from $[\mathbf{h}, \mathbf{s}, \eta, \mathbf{w}]$
evaluate α	from $[\mathbf{R}, \mathbf{Q}]$
evaluate η	from $[\alpha, \mathbf{s}, n, m]$
$\Delta \alpha := \alpha - \alpha_0$;	
$\alpha_0 := \alpha$	
<u>until</u> $ \Delta \alpha \ll \alpha $	

An attractive point of this iteration is that it requires no starting values for α . On the other hand, it requires an initial guess for η , which is much simpler. This property makes it a robust and fast converging algorithm.

When the weighing factors w are *balanced*, meaning that they are all equal for all frequencies (e.g. $w \equiv 1$), the relative error δ is minimized and α is the requested optimal solution. The use of *unbalanced* weighing factors w makes the iterative fit more (or less) sensitive in restricted frequency ranges. In case $w=0$ in a frequency interval, the associated errors are completely ignored. This is handy to exclude spikes and other errors in measurements, or to exclude frequency ranges in which the transfer function is inadequately predicted by the chosen rational function.

Appendix

E. Algorithm for rational magnitude fit

This appendix discusses an algorithm for finding the *real* coefficients of a rational function that optimally fits to a table with measured data, in a special case. The rational function represents a transfer function of the following general form:

$$H(j\omega) = H(s) \stackrel{\text{def}}{=} g \cdot \frac{a_0 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{m-1} \cdot s^{m-1} + s^m}{b_0 + b_1 \cdot s + b_2 \cdot s^2 + \dots + b_{n-1} \cdot s^{n-1} + s^n} = g \cdot \frac{T(s)}{N(s)}$$

The general curve fit is discussed in appendix *D*. This appendix is focused on situations that the information on the transfer function is restricted to $|H(j\omega)|$ or $H(j\omega) / |H(j\omega)|$.

If the information of a rational transfer function $H(j\omega)$ with complex function values is restricted to the real values $|H(j\omega)|$, then the square of these real numbers is a rational function too. Using this property, the coefficients of this related rational function are evaluated as follows:

$$|H(j\omega)|^2 = H(+j\omega) \cdot H(-j\omega) = g^2 \cdot \frac{T(+j\omega)}{N(+j\omega)} \cdot \frac{T(-j\omega)}{N(-j\omega)}$$

$$|H(j\omega)|^2 \stackrel{\text{def}}{=} g^2 \cdot \frac{A_0 + A_2 \cdot s^2 + A_4 \cdot s^4 + \dots + A_{2m-2} \cdot s^{2m-2} + s^{2m}}{B_0 + B_2 \cdot s^2 + B_4 \cdot s^4 + \dots + B_{2n-2} \cdot s^{2n-2} + s^{2n}} \stackrel{\text{def}}{=} G \cdot \frac{T(s^2)}{N(s^2)}$$

$$\beta \stackrel{\text{def}}{=} \begin{bmatrix} G \cdot A_0 \\ G \cdot A_2 \\ \dots \\ G \cdot A_{2m-2} \\ G \\ B_0 \\ B_2 \\ \dots \\ B_{2n-2} \end{bmatrix}$$

$$\beta = \text{TransferFit}(|h|^2, s^2, n, m, \mathbf{w}) \quad \text{see appendix } D \text{ for the definition of function TransferFit}$$

The coefficients in β generate a rational function that describes the function $H(s) \cdot H(-s)$. The poles and zeros of $H(+s)$ are symmetrical situated in the complex plane, with respect to the poles and zeros of $H(-s)$. There are no means available to distinct them in a pole-zero pattern of $H(s) \cdot H(-s)$. Nevertheless, a good choice is based on the assumption that $H(s)$ is a *minimum phase* transfer function.

$H(s)$ is reconstructed from the pole-zero pattern of $H(s) \cdot H(-s)$ by combining all left half-plane singularities and ejecting all other poles and zeros.

Appendix

F. Algorithm for rational delay fit

This appendix discusses an algorithm for finding the *real* coefficients of a rational function that optimally fits to a table with measured data, in a special case. The rational function represents a transfer function of the following general form:

$$H(j\omega) = H(s) \stackrel{\text{def}}{=} \frac{1 - a_1 \cdot s + a_2 \cdot s^2 - \dots - a_{n-1} \cdot s^{n-1} + a_n \cdot s^n}{1 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{n-1} \cdot s^{n-1} + a_n \cdot s^n} = \frac{N(-s)}{N(+s)}$$

when the order n is odd, the last term in the denominator is negative

These rational functions have the property that $|H(j\omega)| \equiv 1$, which makes them attractive to model (pseudo) delay using a *finite* number of poles and zeros. The general curve fit is discussed in appendix *D*. This appendix is a variant of that general fit, and is focused on the differences only.

Let h denote the measured function value of $H(s)$ at complex frequency $s=j\omega$, let w be an arbitrary weighing factor (see appendix *C*), and δ the relative error of h . Then:

$$\frac{1 - a_1 \cdot s + a_2 \cdot s^2 - \dots - a_{n-1} \cdot s^{n-1} + a_n \cdot s^n}{1 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{n-1} \cdot s^{n-1} + a_n \cdot s^n} \approx h$$

$$w \cdot \frac{1 - a_1 \cdot s + a_2 \cdot s^2 - \dots - a_{n-1} \cdot s^{n-1} + a_n \cdot s^n}{1 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{n-1} \cdot s^{n-1} + a_n \cdot s^n} = h \cdot w \cdot (1 + \delta)$$

$$\eta \stackrel{\text{def}}{=} N(s) \stackrel{\text{def}}{=} (1 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{n-1} \cdot s^{n-1} + a_n \cdot s^n)$$

$$w/h \cdot (1 - a_1 \cdot s + a_2 \cdot s^2 - \dots - a_{n-1} \cdot s^{n-1} + a_n \cdot s^n) - w(1 + a_1 \cdot s + a_2 \cdot s^2 + \dots + a_{n-1} \cdot s^{n-1} + a_n \cdot s^n) = w \cdot \eta \cdot \delta$$

$$-w \cdot \{ (1-1/h) + (1+1/h) \cdot a_1 \cdot s + (1-1/h) \cdot a_2 \cdot s^2 + \dots + (1+1/h) \cdot a_{n-1} \cdot s^{n-1} \} = w \cdot \eta \cdot \delta$$

$$(w/\eta) \cdot \{ (1+1/h) \cdot a_1 \cdot s + (1-1/h) \cdot a_2 \cdot s^2 + \dots + (1+1/h) \cdot a_{n-1} \cdot s^{n-1} \} = (w/\eta) \cdot (1/h-1) - w \cdot \delta$$

To simplify these equations, the vectors α , $\mathbf{R}$, $\Delta \mathbf{R}$ and $\mathbf{Q}$ are as follows defined for all frequencies in the table:

$$\alpha \stackrel{\text{def}}{=} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix} \quad \mathbf{R} \stackrel{\text{def}}{=} \begin{bmatrix} (1/h_1-1) \cdot w_1/\eta_1 \\ (1/h_2-1) \cdot w_2/\eta_2 \\ (1/h_3-1) \cdot w_3/\eta_3 \\ \dots \\ (1/h_k-1) \cdot w_k/\eta_k \end{bmatrix} \quad \Delta \mathbf{R} \stackrel{\text{def}}{=} \begin{bmatrix} w_1 \cdot \delta_1 \\ w_2 \cdot \delta_2 \\ w_3 \cdot \delta_3 \\ \dots \\ w_k \cdot \delta_k \end{bmatrix} \quad \mathbf{W} \stackrel{\text{def}}{=} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \dots \\ w_k \end{bmatrix}$$

$$\mathbf{Q} \stackrel{\text{def}}{=} \begin{bmatrix} (1+1/h_1) \cdot w_1/\eta_1 & (1-1/h_1) \cdot w_1/\eta_1 \cdot s_1 & (1+1/h_1) \cdot w_1/\eta_1 \cdot s_1^2 & \dots & (1+1/h_1) \cdot w_1/\eta_1 \cdot s_1^{n-1} \\ (1+1/h_2) \cdot w_2/\eta_2 & (1-1/h_2) \cdot w_2/\eta_2 \cdot s_2 & (1+1/h_2) \cdot w_2/\eta_2 \cdot s_2^2 & \dots & (1+1/h_2) \cdot w_2/\eta_2 \cdot s_2^{n-1} \\ (1+1/h_3) \cdot w_3/\eta_3 & (1-1/h_3) \cdot w_3/\eta_3 \cdot s_3 & (1+1/h_3) \cdot w_3/\eta_3 \cdot s_3^2 & \dots & (1+1/h_3) \cdot w_3/\eta_3 \cdot s_3^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ (1+1/h_k) \cdot w_k/\eta_k & (1-1/h_k) \cdot w_k/\eta_k \cdot s_k & (1+1/h_k) \cdot w_k/\eta_k \cdot s_k^2 & \dots & (1+1/h_k) \cdot w_k/\eta_k \cdot s_k^{n-1} \end{bmatrix}$$

$$\mathbf{Q} \cdot \alpha = \mathbf{R} - \Delta \mathbf{R}$$

$$\alpha = \begin{bmatrix} \mathbf{Q}_{\text{re}} \\ \mathbf{Q}_{\text{im}} \end{bmatrix} \setminus \begin{bmatrix} \mathbf{R}_{\text{re}} \\ \mathbf{R}_{\text{im}} \end{bmatrix}$$

see appendix *B* for the definition of an over-determined left-hand matrix division, restricted to real results

Similarly to the algorithm in appendix *D*, the values of η must be estimated and improved in a similarly iterative way.

Appendix

G. Algorithm for weighed polynomial division

This appendix discusses a generalized algorithm for finding a quotient polynomial of two polynomials. If the rest of a polynomial division is zero, then the quotient polynomial is unique. For non-zero rest polynomials, different quotients exist. Division of $T(s)/N(s)$ results in some $q(s)$ and $r(s)$, that are as follows related to $T(s)$ and $N(s)$:

$$\begin{array}{ll}
 T(s) = q_l(s) \cdot N(s) + r_l(s) & \text{normal (backward) division with remainder} \\
 T(s) = q_f(s) \cdot N(s) + r_f(s) \cdot s^{m-k} & \text{reverse (forward) division with remainder} \\
 T(s) = q_g(s) \cdot N(s) + r_g(s) \cdot W(s) & \text{weighed division with remainder} \\
 \\
 T(s) & = \text{numerator} & q(s) & = \text{quotient} \\
 N(s) & = \text{denominator} & r(s) & = \text{remainder} \\
 W(s) & = \text{weighing function}
 \end{array}$$

The weighing polynomial $W(s)$ specifies what quotient is calculated. This appendix discusses an algorithm to evaluate $\{q(s), r(s)\}$ when $\{T(s), N(s), W(s)\}$ are specified.

When the polynomial coefficients of $q(s)$, $r(s)$ and $T(s)$ are organized as column vectors, and $N(s)$ and $W(s)$ as matrices, then the weighed polynomial quotient can be calculated using matrix division (see appendix B). This is illustrated in the following example:

$$\begin{bmatrix} n0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ n1 & n0 & 0 & 0 & 0 & 0 & 0 & 0 \\ n2 & n1 & n0 & 0 & 0 & 0 & 0 & 0 \\ n3 & n2 & n1 & 0 & 0 & 0 & 0 & 0 \\ n4 & n3 & n2 & 0 & 0 & 0 & 0 & 0 \\ n5 & n4 & n3 & 0 & 0 & 0 & 0 & 0 \\ 0 & n5 & n4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & n5 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} q0 \\ q1 \\ q2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & w0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & w1 & w0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w1 & w0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & w1 & w0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & w1 & w0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & w1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} r0 \\ r1 \\ r2 \\ r3 \\ r4 \end{bmatrix} = \begin{bmatrix} t0 \\ t1 \\ t2 \\ t3 \\ t4 \\ t5 \\ t6 \\ t7 \end{bmatrix}$$

When the individual matrix products are combined, then:

$$\begin{bmatrix} n0 & 0 & 0 & w0 & 0 & 0 & 0 & 0 \\ n1 & n0 & 0 & w1 & w0 & 0 & 0 & 0 \\ n2 & n1 & n0 & 0 & w1 & w0 & 0 & 0 \\ n3 & n2 & n1 & 0 & 0 & w1 & w0 & 0 \\ n4 & n3 & n2 & 0 & 0 & 0 & w1 & w0 \\ n5 & n4 & n3 & 0 & 0 & 0 & 0 & w1 \\ 0 & n5 & n4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & n5 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} q0 \\ q1 \\ q2 \\ r0 \\ r1 \\ r2 \\ r3 \\ r4 \end{bmatrix} = \begin{bmatrix} t0 \\ t1 \\ t2 \\ t3 \\ t4 \\ t5 \\ t6 \\ t7 \end{bmatrix}$$

$$\begin{bmatrix} q0 \\ q1 \\ q2 \\ r0 \\ r1 \\ r2 \\ r3 \\ r4 \end{bmatrix} = \begin{bmatrix} n0 & 0 & 0 & w0 & 0 & 0 & 0 & 0 \\ n1 & n0 & 0 & w1 & w0 & 0 & 0 & 0 \\ n2 & n1 & n0 & 0 & w1 & w0 & 0 & 0 \\ n3 & n2 & n1 & 0 & 0 & w1 & w0 & 0 \\ n4 & n3 & n2 & 0 & 0 & 0 & w1 & w0 \\ n5 & n4 & n3 & 0 & 0 & 0 & 0 & w1 \\ 0 & n5 & n4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & n5 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \setminus \begin{bmatrix} t0 \\ t1 \\ t2 \\ t3 \\ t4 \\ t5 \\ t6 \\ t7 \end{bmatrix}$$

A general division algorithm must copy the polynomial coefficients of $N(s), T(s), W(s)$ to matrices, in accordance with this example, followed by matrix division. The coefficients of the result matrix must be copied to the polynomials $q(s)$ and $r(s)$.

Appendix

H. Algorithm for dominant deflation of transfer order

This appendix discusses an algorithm to approximate a low-pass rational transfer function $H(s)$ with another rational function $H_d(s)$ with lower transfer order. On input, the coefficients of $H(s)$ are provided (or the poles and zeros) and the maximum frequency of interest: ω_0 . On output, the coefficients of $H_d(s)$ are provided by the algorithm. By definition, $H(j\omega_0) = H_d(j\omega_0)$ and $H(j\omega) \approx H_d(j\omega)$ for $\omega < \omega_0$. An application example of this algorithm is loopgain deflation, as discussed in section 4.5

Rational functions are the quotient of two polynomials. Higher order polynomial terms are relevant for 'high' frequencies only, and can be omitted when irrelevant. The consequence of this type of deflation method is that the error of the deflated transfer function increases with the frequency. For loopgain deflation it is crucial that the deflated function value for $\omega = \omega_0$ equals the original function value. Therefore our deflation algorithm is performed in two steps:

- The first step removes all power terms that contribute less than a predefined value, e.g. 30%, for frequencies lower than the bandwidth ($\omega \leq \omega_0$).
- The second step changes the two highest coefficients of numerator as well as denominator to match $H(j\omega_0) = H_d(j\omega_0)$.

We will illustrate this approach using an example of a transfer function with four zeros and five poles. $H(s)$ is deflated into $H_d(s)$ using $H_0(s)$ as intermediate result. To simplify the description, the coefficients of the numerator and the denominator are scaled with respect to ω_0 (the corner frequency of the bandwidth of interest).

$$H(s) = G_0 \cdot \frac{1 + a_1 \cdot (s/\omega_0) + a_2 \cdot (s/\omega_0)^2 + a_3 \cdot (s/\omega_0)^3 + a_4 \cdot (s/\omega_0)^4}{1 + b_1 \cdot (s/\omega_0) + b_2 \cdot (s/\omega_0)^2 + b_3 \cdot (s/\omega_0)^3 + b_4 \cdot (s/\omega_0)^4 + b_5 \cdot (s/\omega_0)^5}$$

$$H_0(s) = G_0 \cdot \frac{1 + a_1 \cdot (s/\omega_0) + a_2 \cdot (s/\omega_0)^2}{1 + b_1 \cdot (s/\omega_0) + b_2 \cdot (s/\omega_0)^2 + b_3 \cdot (s/\omega_0)^3 + b_4 \cdot (s/\omega_0)^4} \quad \text{step 1}$$

$$H_d(s) = G_0 \cdot \frac{1 + \alpha_1 \cdot (s/\omega_0) + \alpha_2 \cdot (s/\omega_0)^2}{1 + b_1 \cdot (s/\omega_0) + b_2 \cdot (s/\omega_0)^2 + \beta_3 \cdot (s/\omega_0)^3 + \beta_4 \cdot (s/\omega_0)^4} \quad \text{step 2}$$

step 1

The removal of power terms in step 1 is an iterative process. The algorithm removes the highest term when it fulfills the requirement that its magnitude at $\omega = \omega_0$ is a factor δ below the magnitude of the summated proceeding terms.

<u>repeat</u>	
<u>if</u> $ a_k < \delta \cdot \text{abs}(1 + a_1 \cdot j + a_2 \cdot j^2 + a_3 \cdot j^3 + \dots + a_{k-1} \cdot j^{k-1})$	<u>then</u> remove a_k
<u>if</u> $ b_k < \delta \cdot \text{abs}(1 + b_1 \cdot j + b_2 \cdot j^2 + b_3 \cdot j^3 + \dots + b_{k-1} \cdot j^{k-1})$	<u>then</u> remove b_k
<u>until</u> nothing to remove	

The iteration proceeds until the above criterion fails.

step 2

The first deflated approximation is exact for $\omega=0$ however the accuracy decreases with increasing frequency. Step two of the deflation is improvement of the deflation accuracy for frequencies near $\omega=\omega_0$. The complex error in the polynomials of $H(s)$ after step 1, and the proposed correction for step 2 is:

error in numerator when deflated from order $M \rightarrow m$ after removal of the associated higher order terms $\delta_a = \sum_{k=m+1}^{k=M} (a_k \cdot j^{(k-m)})$	modification of the highest coefficients in the deflated numerator $\alpha_{m-1} = a_{m-1} - \text{imag}\{\delta_a\}$ $\alpha_m = a_m + \text{real}\{\delta_a\}$
error in denominator when deflated from order $N \rightarrow n$ after removal of the associated higher order terms $\delta_b = \sum_{k=n+1}^{k=N} (b_k \cdot j^{(k-n)})$	modification of the highest coefficients in the deflated denominator $\beta_{n-1} = a_{n-1} - \text{imag}\{\delta_b\}$ $\beta_n = a_n + \text{real}\{\delta_b\}$

In the special case that one of the deflated polynomials is a first order polynomial (or a constant), the proposed algorithm must perform a slightly different approach. It requires the replacement of the correction pair $\{\delta_a, \delta_b\}$ by the factors δ_{a0} or δ_{b0} alone, as defined below:

$$H(j\omega_0) = \frac{T_0(j\omega_0) + \delta_a}{N_0(j\omega_0) + \delta_b} = \frac{T_0(j\omega_0)}{N_0(j\omega_0) + \delta_{b0}} = \frac{T_0(j\omega_0) + \delta_{a0}}{N_0(j\omega_0)}$$

The example in section 4.5 illustrates how important the second step of the algorithm is. In that example, it improves the phase accuracy while the magnitude accuracy was fair.

Appendix

I. Bessel, Butterworth and Chebyshev transfer functions

This appendix provides analytical expressions for standard low-pass transfer functions, normalized with respect to their asymptotic behavior.

The low-pass Bessel, Butterworth and Chebyshev transfer functions are all-pole transfer functions, well-known from filter applications. Many textbook discussions on these functions normalize them therefore with respect to the -3 dB corner frequency. This normalization is inadequate for specifying the generic pass-band in the synthesis of feedback loops (see section 5.1.3 and 5.3.1).

Band-pass synthesis requires normalization with respect to their *asymptotic* behavior. The all-pole transfer functions $H(s)$ in this appendix, with equal transfer order, have equal transfer for $\omega \rightarrow 0$ and for $\omega \rightarrow \infty$. Bessel transfer functions are most conveniently calculated from their polynomial coefficients while the use of polynomial roots is to be preferred for the other functions. They are:

$H(s) = 1/N(s)$	$N(s) = \sum_{k=0}^n (c_k \cdot (s/\omega_0)^k) = \prod_{k=1}^n (1 - (s/\omega_0)/p_k)$
<i>Bessel</i>	$c_k = c_{k-1} \cdot \frac{(n-k+1) \cdot 2q}{(2n-k+1) \cdot k}$ $\left\{ \begin{array}{l} c_0 = 1 \\ q = \sqrt[n]{1 \cdot 3 \cdot 5 \dots (2n-1)} \end{array} \right\}$
<i>Butterworth</i>	$p_k = \exp(j\pi/2 \cdot (2k-1+n)/n)$
<i>Chebyshev</i>	$p_k = \frac{2 \cdot \sinh(\mu/n + j\pi/2 \cdot (2k-1+n)/n)}{ 2 \cdot \cosh(\mu + j\pi/2 \cdot n) ^{1/n}}$ $\left\{ \begin{array}{l} \mu = \operatorname{asinh}(1/\epsilon) \\ \epsilon_{dB} = 10 \cdot \log(1 + \epsilon^2) \end{array} \right\}$

The asymptotic behavior of these functions $N(j\omega)$ is equal for $\omega \rightarrow 0$ or $\omega \rightarrow \infty$, because their first and last polynomial coefficient are all normalized: $c_0 = c_n = 1$.

In addition to the above definitions, simple analytical expressions exist for the denominator coefficients of Butterworth polynomials. They are:

Butterworth polynomial coefficients (exact)							
	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>	
<i>c</i> ₀	1	1	1	1	1	1	<i>c</i> ₀ = 1
<i>c</i> ₁	1	√2	2	√(4+2√2)	1+√5	√2+√6	<i>c</i> ₂ = ½· (<i>c</i> ₁) ²
<i>c</i> ₂		1	2	2+√2	3+√5	4+2·√3	
<i>c</i> ₃			1	√(4+2√2)	3+√5	3√2+2√6	<i>c</i> _k = <i>c</i> _{n-k}
<i>c</i> ₄				1	1+√5	4+2·√3	
<i>c</i> ₅					1	√2+√6	
<i>c</i> ₆						1	

The numerical values of coefficients and poles of all these functions are specified on the next pages.

Bessel polynomial coefficients

	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>
<i>c</i> ₀	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
<i>c</i> ₁	1.0000	1.7321	2.4662	3.2011	3.9363	4.6717
<i>c</i> ₂		1.0000	2.4329	4.3916	6.8864	9.9202
<i>c</i> ₃			1.0000	3.1239	6.7767	12.3583
<i>c</i> ₄				1.0000	3.8107	9.6223
<i>c</i> ₅					1.0000	4.4952
<i>c</i> ₆						1.0000

Butterworth polynomial coefficients

	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>
<i>c</i> ₀	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
<i>c</i> ₁	1.0000	1.4142	2.0000	2.6131	3.2361	3.8637
<i>c</i> ₂		1.0000	2.0000	3.4142	5.2361	7.4641
<i>c</i> ₃			1.0000	2.6131	5.2361	9.1416
<i>c</i> ₄				1.0000	3.2361	7.4641
<i>c</i> ₅					1.0000	3.8637
<i>c</i> ₆						1.0000

Chebyshev polynomial coefficients (1 dB ripple)

	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>
<i>c</i> ₀	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
<i>c</i> ₁	1.0000	1.0455	1.9890	1.9522	3.1074	2.8534
<i>c</i> ₂		1.0000	1.2525	2.7694	3.4289	5.5888
<i>c</i> ₃			1.0000	1.3150	3.9072	4.5796
<i>c</i> ₄				1.0000	1.4249	4.7097
<i>c</i> ₅					1.0000	1.4497
<i>c</i> ₆						1.0000

Chebyshev polynomial coefficients (2 dB ripple)

	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>
<i>c</i> ₀	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
<i>c</i> ₁	1.0000	0.8860	2.1541	1.6916	3.4062	2.4928
<i>c</i> ₂		1.0000	1.0711	2.7699	3.1162	5.5775
<i>c</i> ₃			1.0000	1.0634	4.0834	3.8227
<i>c</i> ₄				1.0000	1.1658	4.6943
<i>c</i> ₅					1.0000	1.1498
<i>c</i> ₆						1.0000

Chebyshev polynomial coefficients (3 dB ripple)

	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>
<i>c</i> ₀	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
<i>c</i> ₁	1.0000	0.7665	2.3356	1.4834	3.7419	2.1967
<i>c</i> ₂		1.0000	0.9473	2.7790	2.8932	5.5884
<i>c</i> ₃			1.0000	0.8967	4.2855	3.2832
<i>c</i> ₄				1.0000	0.9998	4.7014
<i>c</i> ₅					1.0000	0.9596
<i>c</i> ₆						1.0000

Bessel poles angle in degrees

	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>	
p ₁	1.0000	1.0000	1.0305	1.0588	1.0825	1.1022	0.0000	30.0000	43.6525	51.6325	56.9366	60.7508	∠(-p ₁)
p ₂		1.0000	0.9416	0.9444	0.9598	0.9775		-30.0000	0.0000	16.6697	27.4696	35.1079	∠(-p ₂)
p ₃			1.0305	0.9444	0.9264	0.9282			-43.6525	-16.6697	0.0000	11.5411	∠(-p ₃)
p ₄				1.0588	0.9598	0.9282				-51.6325	-27.4696	-11.5411	∠(-p ₄)
p ₅					1.0825	0.9775					-56.9366	-35.1079	∠(-p ₅)
p ₆						1.1022						-60.7508	∠(-p ₆)

Butterworth poles angle in degrees

	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>	
p ₁	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	45.0000	60.0000	67.5000	72.0000	75.0000	∠(-p ₁)
p ₂		1.0000	1.0000	1.0000	1.0000	1.0000		-45.0000	0.0000	22.5000	36.0000	45.0000	∠(-p ₂)
p ₃			1.0000	1.0000	1.0000	1.0000			-60.0000	-22.5000	0.0000	15.0000	∠(-p ₃)
p ₄				1.0000	1.0000	1.0000				-67.5000	-36.0000	-15.0000	∠(-p ₄)
p ₅					1.0000	1.0000					-72.0000	-45.0000	∠(-p ₅)
p ₆						1.0000						-75.0000	∠(-p ₆)

Chebyshev poles (1 dB ripple) angle in degrees

	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>	
p ₁	1.0000	1.0000	1.2636	1.3708	1.5121	1.5545	0.0000	58.4846	75.6524	81.9240	84.8372	86.4183	∠(-p ₁)
p ₂		1.0000	0.6263	0.7295	0.9966	1.1664		-58.4846	0.0000	50.4085	69.0562	76.8514	∠(-p ₂)
p ₃			1.2636	0.7295	0.4403	0.5515			-75.6524	-50.4085	0.0000	48.9176	∠(-p ₃)
p ₄				1.3708	0.9966	0.5515				-81.9240	-69.0562	-48.9176	∠(-p ₄)
p ₅					1.5121	1.1664					-84.8372	-76.8514	∠(-p ₅)
p ₆						1.5545						-86.4183	∠(-p ₆)

Chebyshev poles (2 dB ripple) angle in degrees

	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>	
p ₁	1.0000	1.0000	1.3665	1.4308	1.6102	1.6116	0.0000	63.7041	78.6996	83.7515	86.0357	87.2606	∠(-p ₁)
p ₂		1.0000	0.5355	0.6989	1.0347	1.1971		-63.7041	0.0000	57.4556	73.6398	79.8752	∠(-p ₂)
p ₃			1.3665	0.6989	0.3602	0.5183			-78.6996	-57.4556	0.0000	56.3187	∠(-p ₃)
p ₄				1.4308	1.0347	0.5183				-83.7515	-73.6398	-56.3187	∠(-p ₄)
p ₅					1.6102	1.1971					-86.0357	-79.8752	∠(-p ₅)
p ₆						1.6116						-87.2606	∠(-p ₆)

Chebyshev poles (3 dB ripple) angle in degrees

	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>	<i>n=1</i>	<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>	<i>n=6</i>	
p ₁	1.0000	1.0000	1.4530	1.4651	1.6837	1.6430	0.0000	67.4660	80.6195	84.8580	86.7494	87.7578	∠(-p ₁)
p ₂		1.0000	0.4737	0.6825	1.0685	1.2146		-67.4660	0.0000	62.3240	76.4724	81.6867	∠(-p ₂)
p ₃			1.4530	0.6825	0.3090	0.5011			-80.6195	-62.3240	0.0000	61.3948	∠(-p ₃)
p ₄				1.4651	1.0685	0.5011				-84.8580	-76.4724	-61.3948	∠(-p ₄)
p ₅					1.6837	1.2146					-86.7494	-81.6867	∠(-p ₅)
p ₆						1.6430						-87.7578	∠(-p ₆)

Appendix

J. Algorithm for feedback compensation synthesis

This appendix derives an algorithm that calculates two polynomials required for the compensation synthesis of section 5.3.2. These polynomials represent the modified (profiled) poles and inserted (phantom) zeros of the loop to enable stable feedback operation. The input and output of this compensation algorithm can be denoted as the following function description:

$$[N_x(s), T_x(s)] = \text{CompSynthesis}(g, N_0(s), T_0(s), B_0(s/\omega_0))$$

input/output parameter	description
g	the DC loop gain
$N_0(s)$	A denominator polynomial representing the predefined poles of the loop. These are poles of the uncompensated loop, of which the position remains unchanged when the loop is compensated. By definition: $N_0(0)=1$
$T_0(s)$	A numerator polynomial representing the predefined zeros of the loop. These are zeros of the uncompensated loop, of which the position remains unchanged when the loop is compensated. By definition: $T_0(0)=1$
$B_0(s/\omega_0)$	A numerator polynomial of an all-pole low-pass transfer function. This (normalized) low-pass function specifies the generic passband of the effective-aperture. The frequency ω_0 scales the asymptotic bandwidth of this low-pass function. By definition: $B_0(0)=1$ and $B_0(s) \rightarrow s^{\text{np}}$ for $s \rightarrow j\infty$.
$N_x(s)$	A denominator polynomial representing the modified poles of the loop. These are (profiled) poles of the compensated loop, originating from (profiled) poles of the uncompensated loop. Their new positions facilitate stable operation of the feedback loop. By definition: $N_x(0)=1$
$T_x(s)$	A numerator polynomial representing the inserted (phantom) zeros of the loop. Their positions facilitate stable operation of the feedback loop. By definition: $T_x(0)=1$
compensated loop gain	$\hat{H}(s) = g \cdot \frac{T_0(s) \cdot T_x(s)}{N_0(s) \cdot N_x(s)}$
compensated effective-aperture	$\frac{\hat{H}(s)}{\hat{H}(s)-1} \cdot \frac{1}{T_x(s)} = \frac{g}{g-1} \cdot \frac{T_0(s)}{T_0(s)} \cdot \frac{1}{B_0(s/\omega_0)}$
$T_0(s)$	The polynomial $T_0(s)$ is nearly equal to $T_0(s)$, with the exception of roots in the right complex half-plane (RHP). $T_0(s)$ represents the combination of all LHP-poles, and mirrored RHP poles of $T_0(s)$. The transfer function $T_0(s)/T_0(s)$ is a pseudo delay transfer function with unity magnitude response for all frequencies

The purpose of this appendix is to find poles $N_x(s)$ and zeros $T_s(s)$ facilitating the effective-aperture to meet a predefined low-pass transfer function. The roots of $B(s)$ provide a generic form of this transfer function, however, it requires adequate scaling to

match it with the asymptotic properties of the loopgain. Our compensation algorithm is focused on facilitating the following transfer of the effective-aperture:

$$\underbrace{A_{sa}(s)}_{\substack{\text{system} \\ \text{aperture}}} = \underbrace{\frac{g}{g-1}}_{\substack{\text{nearly unity} \\ \text{gain} \\ \text{magnitude} \\ \text{scaling}}} \cdot \underbrace{\frac{T_0(s)}{T_0(s)}}_{\substack{\text{pseudo} \\ \text{delay} \\ \text{delay} \\ \text{scaling}}} \cdot \underbrace{\frac{1}{B_0(s/\omega_0)}}_{\substack{\text{pass-band} \\ \text{of the loop} \\ \text{bandwidth} \\ \text{scaling}}} = \frac{1}{1-1/\hat{H}(s)} \cdot \frac{1}{T_x(s)}$$

What loopgain $\hat{H}(s)$ is required to facilitate a effective-aperture with this transfer function? To find an answer, the polynomials $N_x(s)$ and $T_x(s)$ will be reconstructed from this relation. Reordering results in:

$$\begin{aligned}
 (1-1/\hat{H}(s)) \cdot T_x(s) &= (1-1/g) \cdot T_0(s)/T_0(s) \cdot B_0(s/\omega_0) \\
 (1-1/\hat{H}(s)) \cdot T_x(s) \cdot T_0(s) &= (1-1/g) \cdot T_0(s) \cdot B_0(s/\omega_0) \\
 T_x(s) \cdot T_0(s) - 1/g \cdot N_x(s) \cdot N_0(s) &= (1-1/g) \cdot T_0(s) \cdot B_0(s/\omega_0) \\
 N_x(s) \cdot N_0(s) - g \cdot T_x(s) \cdot T_0(s) &= (1-g) \cdot T_0(s) \cdot B_0(s/\omega_0) \\
 N_0(s) \cdot (N_x(s)-1) - g \cdot T_0(s) \cdot (T_x(s)-1) &= (1-g) \cdot T_0(s) \cdot B_0(s/\omega_0) - N_0(s) + g \cdot T_0(s)
 \end{aligned}$$

To simplify this equation, we define the polynomials $Q_x(s)$, $R_x(s)$ and $Y_0(s)$ as temporarily substitution quantities.

$$Q_x(s) \stackrel{\text{def}}{=} (1/s) \cdot (N_x(s)-1)$$

$$R_x(s) \stackrel{\text{def}}{=} (1/s) \cdot (T_x(s)-1) \cdot (-g)$$

$$Y_0(s) \stackrel{\text{def}}{=} (1/s) \cdot ((1-g) \cdot T_0(s) \cdot B_0(s/\omega_0) - N_0(s) + g \cdot T_0(s))$$

$$N_0(s) \cdot Q_x(s) + T_0(s) \cdot R_x(s) = Y_0(s)$$

This substitution has reduced the complexity of the equation to a standard format. The unknown polynomials $Q_x(s)$ and $R_x(s)$ result from a weighed polynomial division (see appendix G) of $Y_0(s)$ by $N_0(s)$ using $T_0(s)$ as weighing polynomial.

The formal description of the compensation algorithm is:

$$[N_x(s), T_x(s)] = \text{CompSynthesis}(g, N_0(s), T_0(s), B_0(s/\omega_0))$$

$$T_0(s) = \text{LHP}(T_0) \quad = \text{copy all LHP roots and mirror all RHP roots}$$

$$Y_0(s) = (1/s) \cdot ((1-g) \cdot T_0(s) \cdot B_0(s/\omega_0) - N_0(s) + g \cdot T_0(s))$$

$$[Q_x, R_x] = \text{polydiv}(Y_0, N_0, T_0) \quad = \text{weighed polynomial division with } T_0 \text{ (see appendix G)}$$

$$N_x(s) = 1 + s \cdot Q_x(s)$$

$$T_x(s) = 1 - s \cdot R_x(s)/g$$

Appendix

K. Definitions of spectra and integral transformations

This appendix provides mathematical relations between signal representations in the time domain and the frequency domain; it is the theoretical base for section 7.1.

Spectral methods are powerful method for analyzing harmonic and random signals. The concept of *spectrum* is not uniquely defined. They are usually a modification of the standard Fourier-transform, since the Fourier-transform of (pure) harmonic and random signals does not exist. This is because the Fourier integral of infinite long time intervals is not convergent.

We note that most textbook discussions and publications use the letter S for a quantity named *power spectrum*, *spectral power density*, *spectral density*, *spectral intensity*, *noise power* or *RMS noise* however some of them refer to a single sided spectrum while others meant a double sided spectrum. This indicates the relevance of accurate definitions of what spectrum is used.

<i>averaging</i>			
average, around $t=0$	$\langle q(t) \rangle_T$	$\stackrel{\text{def}}{=} \frac{1}{T} \cdot \int_{-1/2T}^{+1/2T} q(t) \cdot dt$	$T = \begin{matrix} \text{system resolution or} \\ \text{system integration time} \end{matrix}$
average, around $t=t_0$	$\langle q(t) \rangle_{T, t_0}$	$\stackrel{\text{def}}{=} \frac{1}{T} \cdot \int_{t_0 - 1/2T}^{t_0 + 1/2T} q(t) \cdot dt$	
stationairy value	U_0	$\stackrel{\text{def}}{=} \langle U(t) \rangle_\infty$	
average intensity	$(U_{\text{rms}})^2$	$\stackrel{\text{def}}{=} \langle U(t) ^2 \rangle_\infty = \langle U(t) \cdot U^*(t) \rangle_\infty$	
$(u^2(t) = \text{instant intensity})$			
Note that $U^*(t)$ is the complex conjugate of $U(t)$ and that $U'(t)$ is the conjugated transpose ¹ of a signal.			

<i>transformation to the frequency domain</i>			
$j_d^* \{U(t); T\}$	$\stackrel{\text{def}}{=}$	$\frac{1}{T} \cdot \int_{-1/2T}^{+1/2T} U(t) \cdot \exp(-j \cdot 2\pi nt/T) \cdot dt$	(discrete spectrum)
$j_n \{U(t); T\}$	$\stackrel{\text{def}}{=}$	$\frac{1}{\sqrt{1/2T}} \cdot \int_{-1/2T}^{+1/2T} U(t) \cdot \exp(-j \cdot 2\pi ft) \cdot dt$	(noise spectrum)
$j_c \{U(t)\}$	$\stackrel{\text{def}}{=}$	$\int_{-\infty}^{+\infty} U(t) \cdot \exp(-j \cdot 2\pi ft) \cdot dt$	(continuous spectrum)
$D\{U(t)\}$	$\stackrel{\text{def}}{=}$	$\int_{-\infty}^{+\infty} \langle U(t) \cdot U^*(t+\tau) \rangle_\infty \cdot \exp(-j \cdot 2\pi f\tau) \cdot d\tau$	(double sided intensity spec.)
$S\{U(t)\}$	$\stackrel{\text{def}}{=}$	$2 \cdot \int_{-\infty}^{+\infty} \langle U(t) \cdot U^*(t+\tau) \rangle_\infty \cdot \exp(-j \cdot 2\pi f\tau) \cdot d\tau$	(single sided intensity spec.)

¹ Note that $U(t)$ represents a scalar quantity, and not a vector or matrix of scalar quantities. For scalar quantities holds that $U'(t) \equiv U^*(t)$, where $U^*(t)$ represents the complex conjugate of $U(t)$. For matrix quantities $U'(t) \equiv U^{*T}(t)$.

When the interpretation is unambiguous, the following shortcuts may be used:

$$\begin{array}{llll}
 \mathbf{j}_d^* \{U(t); T\} & \rightarrow \tilde{Q}_u(n) & \rightarrow \tilde{U}_n & \tilde{U} \tilde{V} \tilde{I} \tilde{J} \\
 \mathbf{j}_n \{U(t); T\} & \rightarrow N_u(f) & \rightarrow U(f) & U V I J \\
 \mathbf{j}_c \{U(t)\} & \rightarrow F_u(f) & \rightarrow U(f) & U V I J \\
 \mathbf{D} \{U(t)\} & \rightarrow D_u(f) & & \\
 \mathbf{S} \{U(t)\} & \rightarrow S_u(f) & &
 \end{array}$$

The *continuous* Fourier spectrum is only defined for functions that vanish to zero when the time goes to infinity. This Fourier transform is therefore *undefined* for signals such as noise and for periodical signals.

The Fourier *noise* spectrum uses a finite interval T , and is therefore wider applicable. The magnitude of the spectrum increases with the length of the measurement interval, however this effect is compensated by a scaling factor $(1/\sqrt{1/2}T)$ in the special case that random noise is measured. It facilitates that the magnitude of the spectrum, *averaged* over a 'small' frequency band, converges to a finite value when T goes to infinity.

The *discrete* Fourier spectrum is also based on a finite interval, however the associated scaling factor $(1/T)$ is optimized for periodical functions. The magnitude of its spectral components, associated with interval T_1 , equals to that of T_2 when $f(t)=f(t+T_1)=f(t+T_2)$. A Spectrum analyzer indicates a spectrum that is close to the Fourier noise spectrum for random signals, and a spectrum that is close to the discrete Fourier spectrum for periodical signals.

reverse transformation to the time domain

$$\begin{array}{lll}
 U(t) & = & \sum_{n=-\infty}^{+\infty} \tilde{Q}_u(n) \cdot \exp(+j \cdot 2\pi n t / T) \quad \text{periodic function } f(t)=f(t+T) \\
 U(t) & = & \sqrt{1/2} T \cdot \int_{-\infty}^{+\infty} N_u(f) \cdot \exp(+j \cdot 2\pi f t) \cdot df \quad \text{for } (|t| < 1/2 T) \\
 U(t) & = & \int_{-\infty}^{+\infty} F_u(f) \cdot \exp(+j \cdot 2\pi f t) \cdot df
 \end{array}$$

time-domain correlation

$$\begin{array}{llll}
 \text{auto correlation:} & \mathbf{R}\{U(t)\} & \stackrel{\text{def}}{=} R_{uu}(\tau) & \stackrel{\text{def}}{=} \langle U(t) \cdot U^*(t+\tau) \rangle_{\infty} \\
 \text{cross correlation:} & \mathbf{R}\{U(t); V(t)\} & \stackrel{\text{def}}{=} R_{uv}(\tau) & \stackrel{\text{def}}{=} \langle U(t) \cdot V^*(t+\tau) \rangle_{\infty}
 \end{array}$$

$$\begin{array}{ll}
 \langle U(t) \cdot V^*(t+\tau) \rangle_{\infty, t_0} & = \langle U(t) \cdot V^*(t+\tau) \rangle_{\infty, t_0 + \Delta t} \\
 \langle U(t) \cdot V^*(t+\tau) \rangle_{T, t_0} & = \langle U(t-\tau) \cdot V^*(t) \rangle_{T, t_0 + \tau} \\
 \langle U(t) \cdot V^*(t+\tau) \rangle_{\infty} & = \langle U(t-\tau) \cdot V^*(t) \rangle_{\infty}
 \end{array}$$

spectral correlation

$$S_u(f) = 2 \cdot \int_{-\infty}^{+\infty} R_u(\tau) \cdot \exp(-j \cdot 2\pi f \tau) \cdot d\tau = 4 \cdot \int_0^{+\infty} R_u(\tau) \cdot \cos(2\pi f \tau) \cdot d\tau$$

$$D_u(f) = \int_{-\infty}^{+\infty} R_u(\tau) \cdot \exp(-j \cdot 2\pi f \tau) \cdot d\tau = 2 \cdot \int_0^{+\infty} R_u(\tau) \cdot \cos(2\pi f \tau) \cdot d\tau$$

$$R_u(t) = \int_{-\infty}^{+\infty} D_u(f) \cdot \exp(+j \cdot 2\pi f t) \cdot df = \int_0^{+\infty} S_u(f) \cdot \cos(2\pi f t) \cdot df$$

$$\begin{aligned} \mathbf{j}_c \{ \langle U(t) \cdot V^*(t+\tau) \rangle_\infty \} &= \mathbf{j}_c \{ \langle V(t) \cdot U^*(t+\tau) \rangle_\infty \}^* \\ \mathcal{S}\{U(t)\} &= 2 \cdot \mathbf{j}_c \{ \langle U(t) \cdot U^*(t+\tau) \rangle_\infty \} \\ \mathcal{S}\{U(t)+V(t)\} &= \mathcal{S}\{U(t)\} + \mathcal{S}\{V(t)\} + 4 \cdot \text{Real} \{ \mathbf{j}_c \{ \langle U(t) \cdot V^*(t+\tau) \rangle_\infty \} \} \end{aligned}$$

$$\underline{\underline{\mathcal{S}}}\{U(t); V(t)\} \stackrel{\text{def}}{=} \begin{pmatrix} S_{uu}(f) & S_{uv}(f) \\ S_{vu}(f) & S_{vv}(f) \end{pmatrix} \stackrel{\text{def}}{=} \mathbf{j}_c \left\{ \begin{matrix} \langle U(t) \cdot U^*(t+\tau) \rangle_\infty & \langle U(t) \cdot V^*(t+\tau) \rangle_\infty \\ \langle V(t) \cdot U^*(t+\tau) \rangle_\infty & \langle V(t) \cdot V^*(t+\tau) \rangle_\infty \end{matrix} \right\}$$

$$\underline{\underline{\mathcal{S}}}\{J(t); V(t)\} \stackrel{\text{def}}{=} \begin{bmatrix} S_{jl}(f) & S_{jv}(f) \\ S_{vj}(f) & S_{vv}(f) \end{bmatrix} \stackrel{\text{def}}{=} \mathbf{j}_c \left\{ \begin{matrix} \langle J_d(t) \cdot J_d^*(t+\tau) \rangle_\infty & \langle J_d(t) \cdot V_d^*(t+\tau) \rangle_\infty \\ \langle V_d(t) \cdot J_d^*(t+\tau) \rangle_\infty & \langle V_d(t) \cdot V_d^*(t+\tau) \rangle_\infty \end{matrix} \right\}$$

$S_{uu}(f)$ = self-spectral intensity

$S_{uv}(f)$ = cross-spectral intensity

intensity relations (Parseval identities)

$$\begin{array}{l|l} \int_{-\infty}^{+\infty} F_{u1}(f) \cdot F_{u2}^*(f) \cdot df &= \int_{-\infty}^{+\infty} U_1(t) \cdot U_2^*(t) \cdot dt \\ \int_{-\infty}^{+\infty} N_{u1}(f) \cdot N_{u2}^*(f) \cdot df &= \langle U_1(t) \cdot U_2^*(t) \rangle_T \\ \sum_{n=-\infty}^{+\infty} \tilde{Q}_{u1}(n) \cdot \tilde{Q}_{u2}^*(n) &= \langle U_1(t) \cdot U_2^*(t) \rangle_T \end{array} \quad \left| \quad \begin{array}{l} \int_{-\infty}^{+\infty} |F_u(f)|^2 \cdot df = \int_{-\infty}^{+\infty} |U(t)|^2 \cdot dt \\ \int_{-\infty}^{+\infty} |N_u(f)|^2 \cdot df = \langle |U(t)|^2 \rangle_T \\ \sum_{n=-\infty}^{+\infty} |\tilde{Q}_u(n)|^2 = \langle U_1(t) \cdot U_2^*(t) \rangle_T \end{array} \right.$$

$$\langle |U(t)|^2 \rangle_\infty = \langle U(t) \cdot U^*(t) \rangle_\infty = R_u(0) = \int_0^{+\infty} S_u(f) \cdot df = \int_{-\infty}^{+\infty} D_u(f) \cdot df$$

Wiener-Khintchine theorem

This approximation holds for 'sufficient' small resolution bandwidth Δf , and 'sufficient' large measurement interval T . It relates the average spectral magnitude of a noise spectrum to an intensity spectrum:

$$\langle |N_u(f)|^2 \rangle_{\Delta f, f} \approx S_u(f) = 2 \cdot D_u(f)$$

miscellaneous

$$\langle U(t) \rangle_T = N_{u,T}(0) / \sqrt{T} = \tilde{Q}_{u,T}(0)$$

$$\langle |U(t)|^2 \rangle_\infty = R_u(0)$$

$$R_u(+t) = R_u^*(-t) = \text{real}$$

$$S_u(+f) = S_u(-f) = \text{real}$$

$$\tilde{Q}_u(n) = F_u(f_n) \cdot \Delta f \quad \text{when } U(t)=0 \text{ for } (|t| > \frac{1}{2}T) \quad (\Delta f = 1/T; \quad f_n = n \cdot \Delta f)$$

convolution product (smoothing operator)

convolution product

$$(U \otimes V)(\tau) \stackrel{\text{def}}{=} U(t) \otimes V(t) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} U(t) \cdot U(t-\tau) \cdot dt \approx T \cdot \langle U(t) \cdot V(t-\tau) \rangle_T$$

$$\text{commutation:} \quad U \otimes V = V \otimes U$$

$$\text{association:} \quad U \otimes (V \otimes W) = (U \otimes V) \otimes W$$

$$\text{distribution:} \quad U \otimes (V+W) = (U \otimes V) + (U \otimes W)$$

Appendix

L. Algorithm for extraction of equivalent input noise

This appendix describes an algorithm for reconstructing the equivalent excess noise, at the input of an amplifier, relative to well-defined reference planes and source admittances. It requires the detection of the amplifier output noise, while additional noise is supplied to the input using a variable (white) noise source. The source admittance of interest is simulated using an admittance tuner.

At least two output noise levels are required for the extraction. For a description of the application of this algorithm, see section 8.5.3 and Mahmoudi [834]. The basic measurement setup and the various mathematical transformation steps are illustrated in figure 10.1.

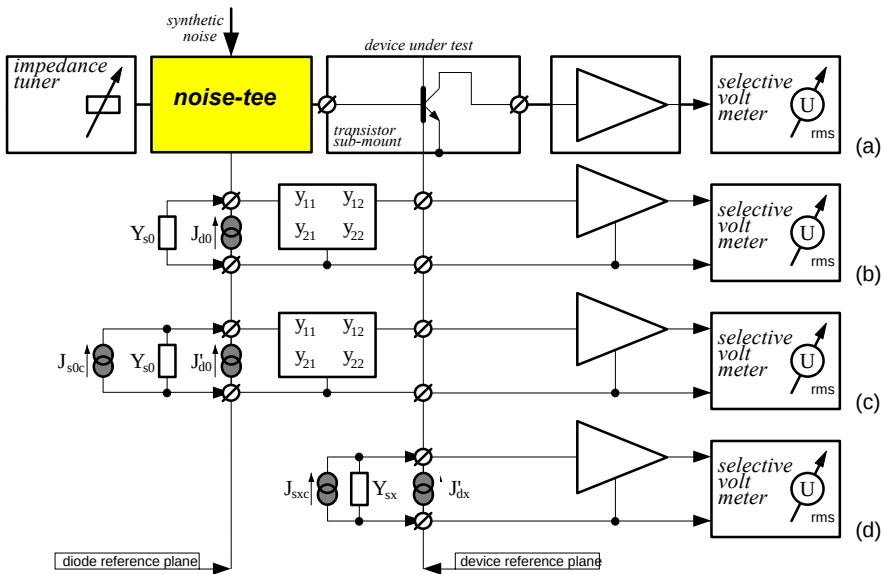


Fig 10.1 Basic measurement setup (a) for the measurement of device noise J_{d0} observed from the diode reference plane for specified admittance Y_{s0} . In (b,c,d) the various steps are illustrated to transform this noise (including thermal noise effects) to the device excess-noise J'_{dx} observed from the device reference plane (excluding thermal noise effects J_{sxc}).

- Figure 10.1a shows a noise-tee. One port is connected with a device under test and the other port is connected with an admittance tuner, to adjust a source admittance of interest. The measurement goal is to extract the equivalent excess-noise, observed from the device reference plane, associated with this source admittance.
- Figure 10.1b shows the measurement of device noise J_{d0} . The noise-tee is illuminated with two or more calibrated synthetic noise levels, to improve overall measurement accuracy. The admittance tuner is adjusted to a specific admittance Y_{s0} of interest. For each reference level S_{rx} , the associated output 'power' P_{rx} is extracted for arbitrary but

sufficient small measurement bandwidth. This 'power' is defined as the square of the detected rms-voltage.

These measurements facilitate the extraction of the equivalent device noise 'observed' from the diode reference plane. The total equivalent noise, 'observed' from this reference plane includes thermal noise effects of tuner admittance and noise-tee losses. The intensity spectrum $S_{J_{d0}}$ of the equivalent device noise current J_{d0} is extracted as follows, using the algorithm of section 8.4.3:

$$\begin{bmatrix} G \cdot S_{J_{d0}} \\ G \cdot S_0 \end{bmatrix} = \begin{bmatrix} 1/P_{r1} & (S_{r1}/S_0)/P_{r1} \\ 1/P_{r2} & (S_{r2}/S_0)/P_{r2} \\ \dots & \dots \\ 1/P_{rn} & (S_{rn}/S_0)/P_{rn} \end{bmatrix} \setminus \begin{bmatrix} 1 \\ 1 \\ \cdot \\ 1 \end{bmatrix} \quad \begin{cases} G = \text{the virtual power 'gain' of the setup} \\ S = \text{an arbitrary scaling factor} \\ S_{rx}^0 = \text{the reference noise levels} \\ P_{rx} = \text{the detected output power} \end{cases}$$

• Figure 10.1c illustrates the extraction of the device *excess-noise* J'_{d0} from J_{d0} and the tuner admittance. In section 7.2.5 we discussed the relation of thermal noise of a passive network with the tuner admittance observed from the diode reference plane. Since this reference plane is not accessible for direct admittance measurement, we reconstruct its value from Y_{sx} using the de-embedding equations of section 2.3.1 and section 8.3.4. Assume, for the time being, that the embedded two-port parameters $[y_{11}, y_{21}, y_{12}, y_{22}]$ and the tuner admittance Y_{sx} are known from other measurements. Then the spectral intensity $S'_{J_{d0}}$ of the excess-noise becomes:

$$S'_{J_{d0}} = S_{J_{d0}} - 4kT \cdot \text{real}(Y_{s0}) = S_{J_{d0}} - 4kT \cdot \text{real} \left\{ \frac{\Delta_y - Y_{sx} \cdot y_{11}}{Y_{sx} - y_{22}} \right\}$$

• Figure 10.1d shows the transformation of the device *excess-noise*, observed from the diode reference plane (J'_{d0}) to the device reference plane (J'_{dx}). Using the de-embedding equations of section 8.3.4 and section 2.3.1 to relate J'_{dx} with J'_{d0} , then we obtain for their intensity spectra:

$$J'_{dx} = \left(\frac{Y_{sx} - y_{22}}{y_{12}} \right) \cdot J'_{d0} \quad \Rightarrow \quad S'_{J_{dx}} = \left| \frac{Y_{sx} - y_{22}}{y_{12}} \right|^2 \cdot S'_{J_{d0}}$$

$$S'_{J_{dx}} = \left| \frac{Y_{sx} - y_{22}}{y_{12}} \right|^2 \cdot \left(S_{J_{d0}} - 4kT \cdot \text{real} \left\{ \frac{\Delta_y - Y_{sx} \cdot y_{11}}{Y_{sx} - y_{22}} \right\} \right) \quad \text{Spectral intensity of device excess noise current.}$$

The thermal noise J_{s0c} is transformed in a similar way. Since the losses in noise-tee and transistor fixture introduce additional thermal noise effects, it is more convenient to combine all these thermal noise effects. This yields an overall extraction of the spectral intensity of the total thermal noise current J_{sxc} , using $S_{J_{sxc}} = \text{real}(Y_{sx})$.

The admittance Y_{sx} is obtained from measurements, using a duplicate of the transistor fixture. Cutting this duplicate fixture into two pieces, gives full access to the (duplicated) device reference plane. It enables admittance measurements using well-known network analyzer techniques.

The embedded two-port parameters $[y_{11}, y_{21}, y_{12}, y_{22}]$ are obtained from full two-port measurements on both ports of the noise-tee, followed by mathematical halving of this two-port using the algorithms of section 8.3.5.

Appendix

M. Noise parameter extraction using hot and cold sources

This appendix describes an algorithm for reconstructing the two-port noise parameters of devices such as transistors. The algorithm requires the measured output noise levels when the device under test (DUT) is connected to a noise source with known source impedance and noise levels, and the DUT input admittance is known. Noise level and noise source admittance are varied to vary the detected output noise levels.

This method is a generalized variant of the methods of Adamian and Uhler [825,830] and Davidson et al. [831], as described in section 8.5.1. Figure 10.2 shows the measurement principle, as described in section 8.5.3, and a linear relation for device excess-noise intensity S_i using $[x_1, x_2, x_3, x_4]$ as noise parameters

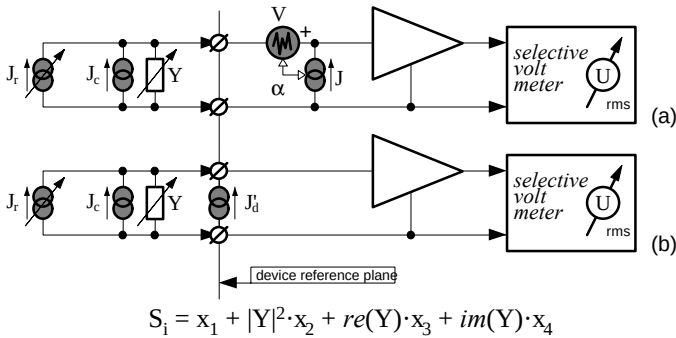


Fig 10.2 The device excess-noise current J'_{d0} with spectral intensity S_p is obtained for arbitrary source admittance values Y from a full noise representation of the input port of the device under test (a). Current J_c is the thermal noise of the tunable source admittance Y and J_r is the tunable source excess noise. The measurement of the noise parameters (c) is performed with a noise source, tunable in reference noise level and in reference source admittance.

The source has admittance Y , (cold) thermal noise current J_c , and an additional reference noise J_r . The noise parameters are extracted from several measurements with these reference noise levels and admittances and from the DUT² input admittance. Since all reference admittances are first to be measured, additional measurement of the input admittance Y_{in} will hardly increase the overall measurement effort. The sensitivity of modern network analyzers enables accurate admittance measurements with low stimulus level to avoid overloading the device under test.

Let P be the detected output noise power (the square of the rms-voltage), when the DUT is connected to a noise source with reference admittance Y . Let $S_{i,tot}$ (or $S_{u,tot}$) be the *total* noise intensity of the equivalent current (or voltage) that is concentrated at the input of the device under test when it is connected to that source. This is the addition of the known (1) thermal noise of the source admittance, the known (2) excess-noise intensity S_r of the noise source, and the unknown (3) excess-noise intensity S_i of the

² DUT = Device under test

DUT. Furthermore, let G be the virtual power gain of the measurement setup, as was defined in section 8.4.1.

Since four unknown noise parameters $[x_1, x_2, x_3, x_4]$ and an additional gain parameter G are to be extracted, at least five noise measurements are required under different circumstances. The detected output power for each noise measurement equals:

$$\begin{aligned}
 P &= G \cdot \{S_{i,\text{tot}}\} \\
 P &= G / |Y + Y_{\text{in}}|^2 \cdot \{S_{i,\text{tot}}\} \\
 P &= G / |Y + Y_{\text{in}}|^2 \cdot \{S_i + S_r + 4 \cdot kT \cdot \text{re}(Y)\} \\
 P &= G / |Y + Y_{\text{in}}|^2 \cdot \{x_1 + |Y|^2 \cdot x_2 + \text{re}(Y) \cdot x_3 + \text{im}(Y) \cdot x_4 + S_r + 4 \cdot kT \cdot \text{re}(Y)\} \\
 \text{let } h &\stackrel{\text{def}}{=} 1 / |Y + Y_{\text{in}}|^2 \\
 \text{let } H &\stackrel{\text{def}}{=} 1 / |Y + Y_{\text{in}}|^2 / P \\
 P &= G \cdot h \cdot \{x_1 + |Y|^2 \cdot x_2 + \text{re}(Y) \cdot x_3 + \text{im}(Y) \cdot x_4 + S_r + 4 \cdot kT \cdot \text{re}(Y)\}
 \end{aligned}$$

The algorithm that we propose extracts the five unknown variables from a set with five or more of these simultaneous equations. Note that there is no serious restriction to the noise source, as long as more than one 'temperature' is used, and four or more different admittance values. One 'hot' measurement and four (or more) passive admittances are adequate. Some multi level measurements (several different 'temperatures' at equal source admittance) are also permitted. Using many levels and admittance values will improve the overall measurement accuracy.

Assume that the relative accuracy of the admittance measurements and of the reference noise levels are superior to the detection errors of the noise power. Then, optimization of the relative error δ in the detected noise power $P \cdot (1 + \delta)$ is equivalent to optimization of the relative accuracy of $S_{i,\text{tot}}$. We assume that this minimization goal provides the most plausible solution for $[x_1, x_2, x_3, x_4, G]$ when the source admittance that will be used in the application is not specified.

To find this solution, we will minimize the *relative* error δ_i . This relative minimization makes the average accuracy of the predicted values S_i insensitive to the admittance Y of interest. Let's consider the following set of simultaneous equations, and solve it with the left-hand matrix division of appendix B:

$$\begin{aligned}
 \begin{bmatrix} P_1(1+\delta_1) \\ P_2(1+\delta_2) \\ \dots \\ \dots \\ P_n(1+\delta_n) \end{bmatrix} &= \begin{bmatrix} h_1 & h_1 \cdot |Y_1|^2 & h_1 \cdot \text{re}(Y_1) & h_1 \cdot \text{im}(Y_1) & h_1 \cdot (S_{r1} + 4kT \cdot \text{re}(Y_1)) \\ h_2 & h_2 \cdot |Y_2|^2 & h_2 \cdot \text{re}(Y_2) & h_2 \cdot \text{im}(Y_2) & h_2 \cdot (S_{r2} + 4kT \cdot \text{re}(Y_2)) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ h_n & h_n \cdot |Y_n|^2 & h_n \cdot \text{re}(Y_n) & h_n \cdot \text{im}(Y_n) & h_n \cdot (S_{rn} + 4kT \cdot \text{re}(Y_n)) \end{bmatrix} \cdot \begin{bmatrix} G \cdot x_1 \\ G \cdot x_2 \\ G \cdot x_3 \\ G \cdot x_4 \\ G \end{bmatrix} \\
 \begin{bmatrix} 1+\delta_1 \\ 1+\delta_2 \\ \dots \\ \dots \\ 1+\delta_n \end{bmatrix} &= \begin{bmatrix} H_1 & H_1 \cdot |Y_1|^2 & H_1 \cdot \text{re}(Y_1) & H_1 \cdot \text{im}(Y_1) & H_1 \cdot (S_{r1} + 4kT \cdot \text{re}(Y_1)) \\ H_2 & H_2 \cdot |Y_2|^2 & H_2 \cdot \text{re}(Y_2) & H_2 \cdot \text{im}(Y_2) & H_2 \cdot (S_{r2} + 4kT \cdot \text{re}(Y_2)) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ H_n & H_n \cdot |Y_n|^2 & H_n \cdot \text{re}(Y_n) & H_n \cdot \text{im}(Y_n) & H_n \cdot (S_{rn} + 4kT \cdot \text{re}(Y_n)) \end{bmatrix} \cdot \begin{bmatrix} G \cdot x_1 \\ G \cdot x_2 \\ G \cdot x_3 \\ G \cdot x_4 \\ G \end{bmatrix}
 \end{aligned}$$

$$\begin{bmatrix} G \cdot x_1 \\ G \cdot x_2 \\ G \cdot x_3 \\ G \cdot x_4 \\ G \end{bmatrix} = \begin{bmatrix} H_1 & H_1 \cdot |Y_1|^2 & H_1 \cdot \text{re}(Y_1) & H_1 \cdot \text{im}(Y_1) & H_1 \cdot (S_{i1} + 4kT \cdot \text{re}(Y_1)) \\ H_2 & H_2 \cdot |Y_2|^2 & H_2 \cdot \text{re}(Y_2) & H_2 \cdot \text{im}(Y_2) & H_2 \cdot (S_{i2} + 4kT \cdot \text{re}(Y_2)) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ H_n & H_n \cdot |Y_n|^2 & H_n \cdot \text{re}(Y_n) & H_n \cdot \text{im}(Y_n) & H_n \cdot (S_{in} + 4kT \cdot \text{re}(Y_n)) \end{bmatrix} \setminus \begin{bmatrix} 1 \\ 1 \\ \dots \\ \dots \\ 1 \end{bmatrix}$$

$$H_x \stackrel{\text{def}}{=} 1 / |Y_x + Y_{in}|^2 / P_x$$

The above expression provides the noise parameters $[x_1, x_2, x_3, x_4]$, without extracting the excess-noise for each source admittance of use.

When all hot sources are matched and all cold sources are arbitrary admittances, then a reference plane transformation is avoided (see sections 8.5.3 and 8.3.4). This can be implemented simply with a noise-tee or a conventional 50Ω noise source. Nevertheless, our generalized algorithm is wider applicable.

When the noise parameters are intended for a wide range of source impedances, choose admittance values that are spread out over the unity reflection circle. For instance: a *load* ($Z=Z_0=50\Omega$), an *open* ($Z=\infty$), a *short* $Z=0$, a *capacitor* ($Z=-j \cdot Z_0$) and an *inductor* ($Z=+j \cdot Z_0$). Otherwise, choose several of them close to the source admittance of the application.

Appendix

N. Noise parameter extraction using paired hot and cold sources

This appendix describes another algorithm for reconstructing the two-port noise parameters of devices such as transistors, similar to the algorithm in appendix *M*. The measurement setup is equal to figure 10.2 and requires a noise source with variable noise level and variable source admittance.

The major difference is that the algorithm requires no information on the input impedance of the device under test. This is an advantage because, the input admittance Y_{in} of the DUT is not always known or available for measurement.

The draw-back of this method, compared to the method of appendix *M* is that measurements must be performed in *pairs*. Each pair is measured with equal source admittance, with two or more noise levels. This is a classic approach (see section 8.5.1, method 2) improved by Lane [823].

At least seven noise measurements are required for defining a set simultaneous equations that has a unique solution for all seven unknown $[x_1, x_2, x_3, x_4, G, re\{Y_{in}\}, im\{Y_{in}\}]$. Unfortunately, this results in a non-linear equation set and will probably require iterative algorithms to find that solution.

When at least four *pairs* of noise measurements are performed, then the equation set can be transformed into a linear form.

The methods of section 8.5.2 are adequate for finding a solution for each pair, associated with admittance Y . When more levels are used in a pair, and when more pairs are used in the overall measurement, then the overall accuracy will improve.

Assume that each pair provides a value for the reconstructed intensity S_i of the device excess-noise current. Furthermore assume that each reconstruction is performed with some relative error δ . As a result, each pair provides the constants for the linear equation:

$$S_i \approx (1+\delta_i) \cdot S_i = x_1 + |Y|^2 \cdot x_2 + re(Y) \cdot x_3 + im(Y) \cdot x_4$$

A set of four or more of these simultaneous equations is adequate to find a solution for the noise parameters $[x_1, x_2, x_3, x_4]$. Let's consider the following set of simultaneous equations, and solve it with the left-hand matrix division of appendix *B*:

$$\begin{bmatrix} 1 & |Y_1|^2 & re(Y_1) & im(Y_1) \\ 1 & |Y_2|^2 & re(Y_2) & im(Y_2) \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 1 & |Y_n|^2 & re(Y_n) & im(Y_n) \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} S_{i1} \cdot (1+\delta_1) \\ S_{i2} \cdot (1+\delta_2) \\ \dots \\ \dots \\ S_{in} \cdot (1+\delta_n) \end{bmatrix}$$

$$\begin{bmatrix} 1/S_{i1} & |Y_1|^2/S_{i1} & re(Y_1)/S_{i1} & im(Y_1)/S_{i1} \\ 1/S_{i2} & |Y_2|^2/S_{i2} & re(Y_2)/S_{i2} & im(Y_2)/S_{i2} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 1/S_{in} & |Y_n|^2/S_{in} & re(Y_n)/S_{in} & im(Y_n)/S_{in} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1+\delta_1 \\ 1+\delta_2 \\ \dots \\ \dots \\ 1+\delta_n \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1/S_{i1} & |Y_1|^2/S_{i1} & re(Y_1)/S_{i1} & im(Y_1)/S_{i1} \\ 1/S_{i2} & |Y_2|^2/S_{i2} & re(Y_2)/S_{i2} & im(Y_2)/S_{i2} \\ \dots & \dots & \dots & \dots \\ 1/S_{in} & |Y_n|^2/S_{in} & re(Y_n)/S_{in} & im(Y_n)/S_{in} \end{bmatrix} \setminus \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$$

The left-hand matrix division minimized the error δ_i in a least-squares sense, which is the *relative* error of the reconstructed input excess-noise current. This relative minimization makes the average accuracy of the predicted values S_i insensitive to the admittance Y of interest.

Appendix

O. Acknowledgments

Support and facilities were essential ingredients for creating this thesis. Therefore, I would like to express my gratitude to the PTT's board of management and the director of this laboratory for their permission to focus on this thesis study for more than two years and to publish it. Within this context, I would also like to thank Henk Groen, Kees van Bochove and Fred Zelders for their management effort in facilitating this study.

Reviews and discussions enabled the overall improvement of this manuscript. Special thanks to Kees de Blok for the initial reviews and the many stimulating discussions on various aspects of this study. I am particularly grateful to Joe Tauritz and Roel Baets of the Delft University of Technology (TUD) for their enthusiasm, discussions and in-depth reviews. Moreover, Joe was of incalculable value for correcting my Dutch variant of 'English' to his native language.

Albert van der Woerd (TUD), Paul Prinz and Hans Schmidt, are acknowledged for reviewing substantial parts of this text and for their valuable suggestions for improvement. Eric Trommel is acknowledged for his stimulating discussions on essentials and applications.

I would like to thank Reza Mahmoudi for his experiments on two-port noise measurements, as part of his graduation project. The fruitful discussions with Johan Kromjong and his calibration activities have contributed to this noise study. I am also grateful to Oskar van Deventer because he facilitated the basis for the noise aspects of this study by constructing the experimental synthetic noise source and by various discussions on this subject. Jan Peter van Bommel is acknowledged for his support on two patent applications, related to this subject.

This thesis study relies on the development of many software elements of the SABEL-CAE system and on the development of our lightwave receivers. I would like to thank all students and colleagues who has contributed to these developments. Special thanks to Paul Prinz, Kees de Blok, Theo van der Vleut, Leon Roos van Raadshooven and Tjeerd van der Laan for their team-spirit on receiver design and SABEL-CAE development. Without the participation of the creators of MatLab and Word-for-Windows this work would not have been completed.

Last but not least, special thanks to Mira for guiding me in finding a fair balance between scientific perfectionist and human being. Her support was of great value to persevere and finish this thesis without losing social contacts completely.

Appendix

P. Bibliography of the author

Rob F.M. van den Brink, was born in Oegstgeest, the Netherlands, in 1955. His route to a university degree was not the most obvious one (MULO, HAVO, HTS). In 1984, he graduated from the Technical University Delft in the Netherlands, with a degree in electrical engineering. His graduation project concerned the development of feeds for microwave antennas.

He joined the same university to work on microwave remote sensing experiments, but changed in 1985 to join PTT Research in the Netherlands. Since then, he worked on fiber-optical transmission systems. He developed wideband optical receivers and transmitters up to a few GHz, and various new measurement and software design tools for opto-electronic circuits. His group designed various analog circuits for coherent optical transmission systems, for internal projects, for an ESPRIT 2054 (UCOL) project supported by the EEC, and later for a RACE 2024 (BAF) project.

These activities formed the technical base for this thesis study. In the period from January 1992 to April 1994, his work was focused on developing the theories and algorithms for this Ph.D.-thesis and on writing this text and various publications.

Two patent applications (pending) resulted from this thesis work.

Appendix

Q. References

Chapter 1: Introduction 100

- [101] R.F.M. van den **Brink**: *Broadband analogue electronics*. PTT Research Yearbook 1989, Leidschendam the Netherlands
- [102] H. de **Ward**: *Optical multigigabit transmission systems*. PTT Research Yearbook 1990, Leidschendam the Netherlands
- [103] T.D. **Poelhekkens**: *Broadband networks for advanced services*. PTT Research Yearbook 1990, Leidschendam the Netherlands
- [104] M.J.M. van **Vaalen**: *Experimental broadband in-house network*. PTT Research Yearbook 1990, Leidschendam the Netherlands
- [105] A.C. **Labrujere**, A.C. van **Bochove**: *Coherent systems*. PTT Research Yearbook 1991, Leidschendam the Netherlands
- [106] R.M. de **Kraker**: *Broadband video networks*. PTT Research Yearbook 1992, Leidschendam the Netherlands
- [107] **Labrujere**, **Deventer**, **Koning**, **Bekooij**, **Tan**, **Roeლოსen**, **Lange**, **Boly**, **Berendschot-Aarts**, **Spruit**, **Nielen**, **Brink**, **Blok**, **Bochove**: *COSNET - a coherent optical subscriber network*. IEEE Journal of lightwave technol, vol 11, no 5/6, pp 865-874, May/June 1993.
- [108] C.M. de **Blok**, R.F.M. van den **Brink**, M.J.M. van **Vaalen**, P.J.M. **Prinz**: *Fast low-cost feed-forward burst-mode optical receiver for 622 Mb/s*: EFOC & N'93 the Eleventh Annual Conference, June 30, 1993. PTT Research, NT-PU-93-1242.
- [109] R.F.M. van den **Brink**: *The characterization and modelling of opto-electronic devices*, Tijdschrift van het NERG, vol 57 no. 3, pp 103-108 (In Dutch), 1992, PTT-Research internal report NT-PU-92-1206.
- [110] R.F.M. van den **Brink**: *Wideband print layout design; analog aspects in designing print layouts for digital circuits*. EMC in theory and practice, Holland Electronica / FME, The Netherlands, June 1993, (in Dutch) ISBN 96-71306-60-7, PTT internal: NT-PU-93-1234
- [111] R.F.M. van den **Brink**: *Handbook SABEL/CAE*. PTT-Research internal report NT-RAP-92-28, Jan 1992 (In Dutch) This is a complete overview of the system and refers to all relevant internal documents
- [112] R.F.M. van den **Brink**: *SABEL/CAE: Microwave toolbox, users manual and theoretical backgrounds version 1.0*. PTT-Research internal report NT-ER-91-1298, Dec 1991 (In Dutch)
- [113] R.F.M. van den **Brink**: *LOCUS Mathematics on some elementary polynomial routines*. PTT-Research internal report 1538DNL/87, 1987 (In Dutch)
- [114] R. **Robbertsen**, R.F.M. van den **Brink**: *LOCUS Mathematics on iteration routines for multi-variable functions (Levenberg-Marquardt iteration)*. PTT-Research internal report 1727DNL/88. (In Dutch)
- [115] R.F.M. van den **Brink**: *LOCUS user's manual version 1.5*, PTT-Research internal report 110RNL/89. 1989 (In Dutch)
- [116] T.A. van der **Laan**, R.F.M. van den **Brink**: *LOCUS algorithm for the determination of the frequency response of electrical circuits, in closed analytical form*. PTT-Research internal report 476RNL/89. 1989, (In Dutch)
- [117] J.J. van **Wamel**, R.F.M. van den **Brink**: *LOCUS a study to automated compensation*. PTT-Research internal report 830RNL/89. 1989 (In Dutch)
- [118] A.J. 't **Jong**, R.F.M. van den **Brink**: *LOCUS Analytical circuit simulation*. PTT-Research internal report 470/91. 1991 (In Dutch)
- [119] C.M.de **Blok**, R.F.M. van den **Brink**: *Full characterization of linear acoustic networks based on N-ports and s-parameters*. J. Audio Eng Soc., Vol 40 no. 6, pp 517-522, June 1992.
- [120] C.M.de **Blok**, R.F.M. van den **Brink**: *Direct-reading one-port acoustic network analyzer*. J. Audio Eng Soc., Vol 41 no. 4, pp 231-238, April 1993.
- [121] R.F.M. van den **Brink**, C.M.de **Blok**: *A one-port acoustic network analyzer with full error correction* - to be published.
- [122] R.F.M. van den **Brink**, C.M.de **Blok**: *Robust calibration methods for one-port acoustic network analyzers with multiple standard reflectors* - to be published.
- [123] **MATLAB**: an interactive mathematical computing environment for numerical computation and visualization. Courtesy of The Mathworks Inc. Natick Mass. USA, FAX: (508)653-2997
- [124] **TOUCHSTONE**: a linear circuit simulator with support of microwave circuit elements, s-parameters, measured data, and optimization. Courtesy of EEsof Inc., Westlake Village, USA.
- [125] **LIBRA**: A non-linear circuit simulator using harmonic balance techniques, compatible with Touchstone. Courtesy of EEsof Inc., Westlake Village, USA.

- [126] **MDS**: A circuit simulation environment (linear and non-linear), support of microwave circuit elements, s-parameters, measured data, and optimization. Courtesy of Hewlett Packard Company, Santa Clara, USA
- [127] **SPICE**: A public domain circuit simulator (linear and non-linear), developed during the mid-1970s. from the University of California, Berkeley, USA.
- [128] **PSPICE**: One of the many commercial derivatives of SPICE. Courtesy of MicroSim Corporation, Irvine, California, USA.

Chapter 2: Formal description of transfer 200

- [201] **F.Strecker, R.Feldtkeller**: *Grundlagen der Theorie des allgemeinen Vierpols*. Elektrische Nachrichten Technik, 6, 93, 1929.
- [202] **L.A.Pipes**: *The matrix theory of four-terminal networks*, Phil Mag. 30, 370, 1940.
- [203] **C.G.Montgomery**: *Techniques of microwave measurements*. MIT Radiation lab serie nr 11, McGraw-Hill Book Co, New York, 1947.
- [204] **C.Montgomery, R.H.Dyke, E.M.Purcell**: *Principles of microwave circuits*. MIT Radiation lab serie nr 8, McGraw-Hill Book Co, New York, 1948.
- [205] **R.F.Shea**, et al.: *Principles of transistor circuits*, John Wiley & Sons inc. New York, 1953.
- [206] **H.J.Carlin**: *The scattering matrix in network theory*, IRE transactions on circuit theory, vol CT-3, pp 88-97, June 1956.
- [207] **P.Penfield**: *Noise in negative resistance amplifiers*, Transactions on circuit theory, Vol CT-7, pp 166-170, 1960.
- [208] **D.C.Youla**: *On scattering matrices normalized to complex port numbers*, Proceedings IRE, vol 49, p 1221, July 1961.
- [209] **P.Penfield**: *A classification of lossless three ports*, Transactions on circuit theory, Vol CT-9, pp 215-223, 1963.
- [210] **K.Kurokawa**: *Power waves and the scattering matrix*. IEEE Transactions on Microwave theory and techniques. Vol MTT-13 no. 2, pp 194-202, March 1965.
- [211] **K.Kurokawa**: *An introduction to the theory of microwave circuits*. Academic press, New york and London 1969.
- [212] **G.A.Deschamps**: *Determination of reflection coefficients and insertion loss of a wave-guide junction*, Journal of applied physics, vol 24 no. 8, pp 1046-1050, Aug 1953.
- [213] **R.W.Anderson, O.T.Dennison**: *An advanced new network analyzer for sweep-measuring amplitude and phase from 0.1 to 12.4 GHz*, Hewlett-Packard Journal, vol 18 no. 6, Feb 1967.
- [214] **R.A.Hackborn**: *An automatic network analyzer system*. The microwave journal, pp45-52, May 1968.
- [215] **IEC/CEI**: *Letter symbols, including conventions and signs for electrical technology*. NNI (ordernr SPE 90079), Delft, The Netherlands 1985
- [216] **I.V. Lindell**: *Methods for electromagnetic field analysis*, Clarendon Press, Oxford 1992.

Chapter 3: Extraction of transfer parameters 300

- [301] **G.Gonzalez**: *Microwave transistor amplifiers*. Prentice Hall, Inc., London 1984.
- [302] **J.J. Ebers, J.L.Moll**: *Large signal behavior of junction transistors*, Proc IRE, vol 42, 1954.
- [303] **H.K.Gummel, H.C.Poon**: *An integral charge control model of bipolar transistors*. Bell System Techn, Journal, vol 49, 1970.
- [304] **I. Getreu**: *Modeling the bipolar transistor*. Tektronix Inc, Beaverton Or., 1976
- [305] **H.C.de Graaf, F.M.Klaasen**: *Compact transistor modelling for circuit design*. (MEXTRAM-model), Springer-Verlag, Wien, 1990

Chapter 4: Formal description of feedback 400

- [401] **H.S. Black**: *Stabilized Feedback Amplifiers*, Bell system Tech. J., 1934.
- [402] **H.S. Black**: *U.S. patent 2,102,671* 1934.
- [403] **H.W. Bode**: *Network Analysis and Feedback Amplifier Design*, van Nostrand Company, Inc, Princeton, N.J., 1945.
- [404] **M.S. Ghauri, D.O. Pederson**: *A new design approach for feedback amplifiers*, Trans.Inst. Radio Engrs., VT-8, 274, September 1961. (probably the first explicit statement of phantom compensation)
- [405] **E.M. Cherry, D.E. Hooper**, : *Amplifying Devices and Low-Pas Amplifier Design*, J Wiley and Sons, Inc, New York, 1968.
- [406] **E.H. Nordholt**: *Design of High-Performance Negative-Feedback Amplifiers*, Elsevier Scientific publ. Company, Amsterdam, 1983.

- [407] P.W. **Tuinenga**: *A guide to circuit simulation and analysis using PSpice*, Prentice Hall, Englewood Cliffs, New Jersey, 1988.
- [408] R. **Molich-Pedersen**: *User's manual for the ANP3 circuit analysis program*, Laboratory of Circuit Theory, Technical University of Denmark, Lyngby, March 1972.
- [409] S. **Skelboe**: *A universal formulae for network functions*, IEEE trans. on Circuits and systems, pp 58-60, Jan 1975.
- [410] D. **Marquardt**: *An algorithm for least-squares estimation of non-linear parameters*, SIAM J. Appl Math, vol 11, pp 431-441, 1963.
- [411] G.C. **Temes**, D.A. **Calahan**: *Computer-aided network optimization, the state of the art*, Proc IEEE vol 55, pp 1832-1863, 1967.
- [412] J.W. **Bandler**, S.H. **Chen**: *Circuit optimization: The state of the art*, IEEE trans. Microwave theory and techn., vol 36 no. 2, pp 424-443, Feb 1988.
- [413] W.H. **Press**, B.P. **Flannery**, S.A. **Teukolsky**, W.T. **Vetterling**: *Numerical recipes in Pasca,l* Cambridge university press, Cambridge 1989.
- [414] L.W. **Johnson**, R.D. **Ries**, J.T. **Arnold**: *Introduction to linear algebra*, Addison-Wesley Publishing Company, Reading, Massachusetts 1989.

Chapter 5: Wideband feedback synthesis 500

- [501] M.G.R.**Degrauwe**, W.M.C.**Sansen**: *A synthesis program for operational amplifiers*, IEEE International Solid-State circuits conference. Digest of technical papers, San Francisco, CA/USA, 22-24, pp 18-19, Feb 1984.
- [502] M.G.R.**Degrauwe**, O.**Nys**, E.**Dijkstra**: *IDAC: An interactive design tool for analog CMOS circuits*, IEEE J. of solid-state circuits, 22 no. 6, pp 1106-1116, Dec 1987.
- [503] F. **El-Turky**, E.E. **Perry**: *BLADES: An artificial intelligence approach to analog circuit design*, IEEE Transactions. on CAD, vol 8 no. 6, pp 680-692, June 1989.
- [504] R. **Harjany**, R.A. **Rutenbar**, L.R.**Carley**: *OASYS: A framework for analog circuit synthesis*. IEEE Transactions. on CAD, vol 8 no. 12, pp 1247-1266, Dec 1989.
- [505] H.Y. **Koh**, C.H. **Séquin**, P.R. **Gray**: *OPASYN: A compiler for CMOS operational amplifiers*, IEEE Transactions. on CAD, vol 9 no. 2, pp 113-125, Feb 1990.
- [506] Z.Q. **Ning**, T. **Mouthaan**, H.**Wallinga**: *SEAS: A simulated evolution approach for analog circuit synthesis*, In CICC'90, May 1990.
- [507] H. **Onodera**, H. **Kanbara**, K. **Tamaru**: *OAC: Operational amplifier compilation with performance optimization*. IEEE J. of solid-state circuits, Vol 25 no. 2, pp 466-473, April 1990.
- [508] W.A.J. **Waller**, M.H.**Smith**, L.T.**Walczowski**, D.**Howard**: *ODIN: A software system for synthesizing CMOS op-amps*. Proceedings of the concurrent Engineering and Design Automation conference, pp 74-77, Bournemouth Polytechnic, Woodhead Publ. Ltd., March 1991.
- [509] J. **Stoffels**: *Automated frequency compensation in high-performance amplifiers*, 30th Midwest symposium on Circuits and Systems, Elsevier Science Publishing Co. Inc., pp 223-226, 1988.
- [510] J. **Stoffels**, C. **Reeuwijk**: *AMPDES: a program for the synthesis of High-Performance Amplifiers*, Proceedings of the european conference on design automation, Brussels, pp 474-479, March 1992.
- [511] G. **Gielen**, W. **Sansen**: *Symbolic analysis for automated design of analog integrated circuit*. Kluwer, Dordrecht, The Netherlands 1991.
- [512] K. **Swings**, W. **Sansen**: *ARIADNE: a constrained-based approach to computer-aided synthesis and modeling of analog integrated circuits*. Analog integrated circuits and signal processing, vol 3, pp 197-215, 1993.
- [513] G. **Gielen**, P. **Wambacq**, W. **Sansen**: *Symbolic analysis methods and applications for analog circuits*. Proceedings of the IEEE, vol 82, no. 2, pp 287-304, February 1994.

Chapter 6: Design example: Lightwave receivers 600

- [601] J.L.**Hullett**, T.V.**Muoi**: *A modified receiver for digital optical fiber transmission systems*. IEEE Trans. Commun, COM 23, pp 1517-1521, 1975.
- [602] J.L.**Hullett**, T.V.**Muoi**: *A feedback receiver amplifier for optical transmission systems*. IEEE Trans. Commun, COM-24, pp 1180-1185, 1976.
- [603] J.X.**Kan**, G.**Jacobsen**, E.**Bodker**: *Noise performance of Gbit/s tuned optical receivers*. Electronic Letters, Vol 23 no. 9, pp 434-435, April 1987.
- [604] G.**Jacobsen**, J.X.**Kan**, I.**Garrett**: *Tuned front-end design for heterodyne optical receivers* J. of Lichtwave technology, vol 7, no 1, pp 105-114, Jan 1989.

- [815] **Hurm, Ludwig, Rosenzweig, Benz, Berroth, Bosch, Bronner, Hülsmann, Köhler, Raynor, Schneider:** *14 GHz bandwidth MSM photodiode AlGaAs/GaAs HEMT monolithic integrated receiver*, El Letters, vol 29 no. 1, pp 9-10, Jan 1993.
- [606] R.F.M. van den **Brink:** *Noise Optimum of a fibre optic receiver pre-amplifier* PTT-Research, Memorandum 1391 DNL/86, Dec 1986
- [607] M.J. van der **Pol**, R.F.M. van den **Brink:** *A wideband optical receiver using third order current-current feedback* (In Dutch), PTT-Research, Memorandum 1526 DNL/87, Juli 1987.
- [608] R.F.M. van den **Brink:** *Optical receiver with third-order capacitive current-current feedback*, Electronics Letters, vol 24 no. 16, pp 1024-1025, Aug 1988.
- [609] P.J.M. **Prinz**, R.F.M. van den **Brink:** *Report of first prototype of 900 MHz optical receiver preamplifier, deliverable 4.23*, PTT-Research, International ESPRIT 2054: UCOL report 938/90, okt 1990.
- [610] R.F.M. van den **Brink:** *Optical receivers with capacitive current-current feedback*, ESPRIT Conference/exhibition, PTT-Research, International ESPRIT 2054: UCOL report 1139/90, Nov 1990.
- [611] H.J.T. van der **Vleut:** *Tuning techniques for improving the noise performance of lightwave receivers.* (In Dutch), PTT-Research, Deelrapport 159/91, March 1991.
- [612] P.J.M. **Prinz**, R.F.M. van den **Brink:** *Report of prototype of receiver preamplifier without tuning, deliverable 4.24*, PTT-Research, International ESPRIT 2054: UCOL report 483/91, April 1991.
- [613] E. **Drijver**, R.F.M. van den **Brink**, H.J.T. van der **Vleut**, P.J.M. **Prinz:** *Report on Tuning, deliverable 4.39*, PTT-Research, International ESPRIT 2054: UCOL report NT-INT-91-1456, Dec 1991.

Chapter 7: Formal description of noise 700

- [701] W.**Schottky**, Ann. Physik, vol 57, p 541, 1918.
- [702] H.**Nyquist:** *Thermal agitation of electric charge in conductors*. Phys Review, vol 32, pp 110-113, July 1928.
- [703] N. **Wiener:** *Generalized harmonic analysis*, Acta Mathematica, Vol 55, pp 117-258, 1930.
- [704] A. **Khinchine:** Math. Ann., vol 109, p 604, 1934.
- [705] R.E. **Burgess:** *Noise in receiving aerial systems*. Proc Phys soc. vol 53, pp293-304, 1941.
- [706] H.T. **Friis:** *Noise figures of radio receivers*, Proceedings IRE, vol 32, pp 419-422, July 1944.
- [707] D.O.**North:** *Discussion of [706]*, Proc IRE, vol 33, pp 125-127, 1945.
- [708] H.**Goldberg:** *some notes on noise figures*, Proc. IRE, vol 36, p 1205, 1948.
- [709] N. **Wiener:** *Extrapolation, Interpolation and smoothing of stationary time systems*, John Wiley & Sons, Inc., New York, N.Y. 1949.
- [710] A. van der **Ziel:** *Noise*. Prentice Hall, New York 1954.
- [711] R.Q. **Twiss:** *Nyquist and Thévenin's theorem generalized for nonreciprocal linear networks*. Journal of applied Physics, vol 26, pp 599-602, May 1955.
- [712] H.**Rothe**, W.**Dahlke:** ---, Archiv der Elektrischen Übertragung, vol 9, pp 117-121, 1955. (In German)
- [713] H.**Rothe**, W.**Dahlke:** *Theory of noisy fourpoles*, Proceedings of the IRE, Vol 44, pp 811-818, June 1956.
- [714] H.**Bauer**, H.**Rothe:** *Der äquivalente rauschvierpol als wellenvierpol*. Archiv der Elektrischen Übertragung, vol 10, pp 241-252, June 1956. (In German)
- [715] H.A.**Haus**, R.B.**Adler:** *Optimum Noise performance of linear amplifiers*, Proceedings of the IRE, Vol 46, pp 1517-1533, 1958.
- [716] H.A.**Haus**, R.B.**Adler:** *Circuit theory of linear noisy networks*. MIT technology press & John Wiley and Sons, New York, 1959.
- [717] H. **Gamo:** *On passive one-way systems*. (Int. symposium on circuit and information theory), IEEE Transactions on circuits and systems, vol CT-6, special supplement, pp 283-298, May 1959.
- [718] P.**Penfield:** *Wave representation of amplifier noise*, Transactions on circuit theory, pp 84-86, 1962.
- [719] H. **Bosma:** *On the theory of linear noisy systems*, Thesis, Techn. University Eindhoven, The Netherlands, 1967 (also Philips research reports suppl, no 10, 1967).
- [720] A van der **Ziel:** *Noise; Sources, Characterization, Measurement*, Prentice-Hall, Englewood Cliffs, New Jersey, 1970.
- [721] H.**Hillbrand**, P.H.**Russer:** *An efficient method for computer aided noise analysis of linear amplifier networks*, IEEE Transactions on circuits and systems, vol CAS-23, pp 235-238, April 1976.
- [722] K.**Hartmann:** *Noise characterization of linear circuits*, IEEE Transactions on circuits and systems, vol CAS-23 no. 10, pp 581-590, Oct 1976.
- [723] V.**Rizzoli**, A.**Lipparini:** *Computer-aided noise analysis of linear multiport networks of arbitrary topology*. IEEE Trans. on Microwave theory and techniques, vol MTT-33 no. 12, Dec 1985.

- [724] J.A. **Dobrowolski**: *Introduction to computer methods for microwave circuit analysis and design*. Artech House, Boston, London, 1991
- [725] S.W. **Wedge**, D.B. **Rutledge**: *Noise waves in passive linear multi-ports*. IEEE Microwave and guided wave letters, vol 1 no. 5, pp 117-119, May 1991.
- Chapter 8: Lightwave based electrical noise measurements** 800
- [801] R.F.M. van den **Brink**, E. **Drijver**, M.O van **Deventer**: *Novel noise measurement setup with high dynamic range for optical receivers*, Electronics Letters, vol 28 no. 7, pp 629-630, March 1992.
- [802] R.F.M. van den **Brink**: *Novel electrical noise source based on lightwave components*, Microwave and Optical technology letters, vol 5 no. 11, pp 549-553, Oct 1992.
- [803] R.F.M. van den **Brink**: *Method for lightwave noise measurements*. Patent application, Europe: EP-0522614, USA, June 1991.
- [804] **Hewlett Packard** *Application note AN 150, Spectrum Analysis*, pp 31-33.
- [805] B.L. **Kasper**, C.A. **Burrus**, J.R. **Talman**, K.L. **Hall**: *Balanced dual-detector receiver for optical heterodyne communication at Gbit/s rates*. Electronics Letters, vol 22 no. 8, pp 413-414, April 1986.
- [806] S. **Machida**, Y. **Yamamoto**: *Quantum-limited operation of balanced mixer homodyne and heterodyne receivers*, IEEE Journal of Quantum Electronics, Vol QE-22 no. 5, pp 617-624, May 1986.
- [807] K. **Ogawa**, E.L. **Chinnoek**: *GaAs FET Transimpedance front-end design for a wideband optical receiver*, Electronics Letters, Vol 15 no. 20, pp 650-652, Sept 1979.
- [808] M.A.M. **Barros**: *Low-noise InSb Photodetector preamp for the infrared*, IEEE Journal of solid-state circuits, vol sc-17 no. 4, pp 761-766, Aug 1982.
- [809] K. **Yamashita**, T. **Kinoshita**, M. **Maeda**, K. **Nakazato**: *Simple common-collector full-monolithic preamplifier for 560 Mbit/s optical transmission*, Electronics Letters, Vol 22 no. 3, pp 146-147, Jan 1986.
- [810] J.L. **Gimlett**: *Low-noise 8GHz PIN/FET optical receiver*, Electronics Letters, Vol 23 no. 6, pp 281-283, March 1987.
- [811] T.E. **Darcie**, B.L. **Kasper**, J.R. **Talman**, C.A. **Burrus**: *Resonant PIN-FET receivers for lightwave subcarrier systems*, IEEE Journal of lightwave technology, vol 6 no. 4, pp 582-589, April 1988.
- [812] S.R. **Cochran**: *Low-noise receivers for fibre-optic microwave signal transmission*, IEEE Journal of lightwave technology, vol 6 no. 8, pp 1328-1337, Aug 1988.
- [813] W.S. **Lee**, D.A.H. **Spear**, M.J. **Agnew**, P.J.G. **Dawe**, S.W. **Bland**: *1.2Gbit/s fully integrated transimpedance optical receiver OEIC for 1.3-1.55 μ m transmission systems*. Electronics Letters, Vol 26 no. 6, pp 377-379, March 1990.
- [814] T. **Olson**: *2.2GHz dual-detector lightwave receiver employing current shunt feedback amplifier design*. Electronics Letters, Vol 26 no. 8, pp 510-512, April 1990.
- [815] V. **Hurm**, et al.: *14GHz bandwidth MSM photodiode AlGaAs/GaAs HEMT monolithic integrated optoelectronic receiver*, Electronics Letters, Vol 29 no. 1, pp 9-10, Jan 93.
- [816] E. **Eichen**, J. **Schlafer**, W. **Rideout**, J. **McCabe**: *Wide bandwidth receiver / photodetector frequency response measurements using amplified spontaneous emission from a semiconductor optical amplifier*, IEEE Journal of lightwave technology, Vol 8 no. 6, pp 912-915, June 1990.
- [817] **IRE subcommittee on noise**, A.G. **Jensen**, et al.: *IRE standards on electron devices: methods of measuring noise*. Proc. of the IRE, Vol 41, pp 890-896, July 1953.
- [818] **IRE subcommittee on noise**, H.A. **Haus**, et al.: *IRE standards on methods of measuring noise in linear twoports*. Proceed. of the IRE, Vol 48, pp 60-6, Jan 19608.
- [819] **IRE subcommittee on noise**, H.A. **Haus**, et al.: *Representation of noise in linear twoports*. Proceedings of the IRE, Vol 48, pp 69-74, Jan 1960.
- [820] **IRE subcommittee on noise**, H.A. **Haus**, et al.: *IRE Standards on electron tubes: Definition on terms*. Proceed. of the IEEE, Vol 51, Jan 1960, pp 434-435, March 1963.
- [821] **IRE subcommittee on noise**, H.A. **Haus**, et al.: *Description of the noise performance of amplifiers and receiving system*. Proceedings of the IEEE, Vol 51, pp 436-442, March 1963.
- [822] J. **Lange**: *Noise characterization of linear twoports in terms of invariant parameters*, Journal of solid state circuits, Vol sc-2 no. 2, pp 37-40, June 1967.
- [823] R.Q. **Lane**: *The determination of device noise parameters*, Proceedings of the IEEE, Vol 116, pp 1461-1462, Aug 1969.
- [824] G.F. **Engen**: *A new method of characterizing amplifier noise performance*, IEEE Transact. on instrumentation and measurement, vol IM-19 no. 4, pp 344-349, Nov 1970.
- [825] V. **Adamian**, A. **Uhlir**: *A novel procedure for receiver noise characterization*, IEEE trans. on instrumentation and measurement, Vol IM-22, pp 181-182, June 1973.

- [826] R.P.Meys: *A wave approach to the noise properties of linear microwave devices*, IEEE trans. on microwave theory and techniques, vol MTT-26 no. 1, pp 34-37, Jan 1978.
- [827] G.Caruso, M.Sannino: *Computer-aided determination of microwave two-port noise parameters*. IEEE trans. on microwave theory and techniques, vol MTT-26 no. 9, pp 639-642, Sept 1978
- [828] M.Mitama, H.Katoh: *An improved computational method for noise parameter measurement*. IEEE Trans. on microwave theory and techniques, vol MTT-27 no. 6, pp 612-615, June 1979.
- [829] M.S.Gupta: *Impossibility of linear two-port noise parameter measurement with a single temperature noise source*. IEEE trans. on instrumentation and measurement, Vol IM-32 no. 3, pp 443-445, Sept 1983.
- [830] V.Adamian, A.Uhlir: *Simplified noise evaluation of microwave receivers*, IEEE trans. on instrumentation and measurement, Vol IM-33 no. 2, pp 136-140, June 1984.
- [831] A.C.Davidson, B.W.Leake, E.Strid: *Accuracy improvements in microwave noise parameter measurements*, IEEE Trans. on microwave theory and techniques, vol MTT-37 no. 12, pp 1973-1977, Dec 1989. See also: A. Uhlir: *Comments*, Jan. 1991, p 157
- [832] J.M.O'Callaghan, J.P.Mondal: *A vector approach for noise parameter fitting and selection of source admittances*, IEEE Trans. on microwave theory and techniques, vol MTT-39 no. 8, pp 1376-1382, Aug 1991.
- [833] S.W.Wedge, D.B.Rutledge: *Wave techniques for noise modeling and measurement*. IEEE Trans. on microwave theory and techniques, vol MTT-40 no. 11, pp 2004-2012, Nov 1992.
- [834] R.Mahmoudi: *Measurement setup for transistor noise parameters*, graduate project, mentored by R.F.M. van den Brink; PTT-Research internal report: NT-SV-93-612, University Delft internal report: 68-220-A194-93, Aug 1993. (In Dutch)
- [835] C.Hentschel: *Fiber optics handbook*, Hewlett-Packard, Boeblingen Instruments Division, Germany, third edition, March 1989.

Chapter 9: Lightwave synthetic noise generation 900

- [901] R.F.M. van den Brink: *Lightwave noise source*. Patent application, The Netherlands: NL-9300347, Feb 1993.
- [902] R.F.M. van den Brink: *Improved lightwave synthetic noise generator using noise injection and triangular modulation*. IEEE Photonics technology letters, vol 6, no 4, pp579-582, April 1994.
- [903] J.A. Armstrong: *Theory of interferometric analysis of laser phase noise*, J. Optical Society of America, vol 56, pp 1024-1031, July 1966.
- [904] T.Okoshi, K.Kikuchi, A.Nakayama: *Novel method for high resolution measurement of laser output spectrum*, Electr. Letters, vol 16 no. 16, pp 630-631, July 1980.
- [905] Y.Yamamoto, T.Mukai, S.Saito: *Quantum phase noise and linewidth of a semiconductor laser*, Electronics Letters, vol 17 no. 9, pp 327-329, April 1981.
- [906] C.H.Henry: *Theory of the linewidth of semiconductor lasers*, IEEE J. of Quantum Electronics, vol. QE-18(2), pp 259-264, February 1982.
- [907] J.Wang, U.Krüger, B.Schwartz, K.Petermann: *Measurement of frequency response of photoreceivers using self-homodyne method*, Electronics Letters, vol 25 no. 11, pp 722-723, May 1989.
- [908] M.O van Deventer: *Self-homodyne sweeper for characterization of optical receivers*, PTT-Research internal report NT383/90, July 1990. (In Dutch)
- [909] D.R. Huber: *Broad linewidth lasers for optical fiber communication systems*. European patent application EP.0.503.579.A2, filed 11-03-1992, published 16-09-92 bulletin 92/38
- [910] A. Yariv: *Optical electronics*, (edition 4), Saunders College Publishing, Philadelphia, 1991.
- [911] M.R. Spiegel: *Mathematical handbook*, Schaum's outline series in mathematics, McGraw-Hill, New York, 1968.

Appendix

R. Index

A

a-parameters • 2-18, 4-19, 7-18, @-1
 ABCD-parameters – See a-parameters • 2-18
 active decoupling • 1-7, 1-10, 5-1, 6-7, 6-8
 cascade stage • 4-25
 long tailed pair • 4-25
 admittance
 – See also y-parameters • 2-13
 admittance matrix • 2-13
 admittance tuner – See measurements (of noise) • 8-10
 load admittance • 2-6
 source admittance • 2-7
 algorithms
 – See also curve fitting • 3-2
 – See also extraction • 3-4
 bandwidth algorithm • 5-8
 circuit noise analysis • 7-13
 circuit transfer analysis • 2-16
 compensation algorithm • 5-25, @-16
 interconnecting waves • 2-8
 loop gain deflation • @-11
 dominant approximation • 4-32
 pole-zero cancellation • 4-31
 using parasitic delay • 4-29
 mathematical halving of noise-tees • 8-23
 matrix division • @-2
 polynomial division • 4-31, @-10
 polynomial polishing • 4-31
 reduction of ports • 2-15, 7-12
 transformation of noise parameters • 7-10, 7-22, 7-25
 transformation of noise through reference planes • 8-22
 transformation of transfer parameters • 2-12, 2-19, @-1
 aperture • 5-3
 – See also available bandwidth • 5-9
 – See also compensation • 5-3
 – See also feedback synthesis • 5-3
 asymptotic approximations • 5-7
 bandpass analysis • 5-9
 bandpass synthesis • 5-22
 effective-aperture • 5-4, 5-18
 generic passband • 5-10, 5-23
 specific passband • 5-23, 5-25
 virtual-aperture • 5-4
 asymptotic corner frequency • 5-8, 5-9, 5-10, @-13
 asymptotic gain model – See feedback model • 4-3
 asymptotic system gain – See superposition parameters • 4-7
 auto correlation • 7-4, @-18
 autonomous noise parameters – See noise parameters • 7-24

available bandwidth • 5-6
 – See also aperture • 5-9
 bandpass analysis • 5-9
 bandpass synthesis • 5-22
 bandwidth algorithm • 5-8
 GB-product • 5-6
 LP-product • 5-6
 average power flow – See also power • 2-5

B

b-parameters • 2-18, 7-18
 bandpass analysis/synthesis
 – See aperture • 5-9
 – See available bandwidth • 5-9
 bandwidth
 – See available bandwidth • 5-6
 – See resolution bandwidth • 7-5
 base resistance – See BJT • 3-18
 Bessel response • 5-10, @-13
 BJT
 base resistance • 3-18
 current gain • 2-23, 3-16
 current transport concept • 3-21
 current transport factor • 2-23, 3-19
 diffusion capacitance • 3-20
 emitter inductance • 3-14, 3-18
 parameter definition • 2-23
 parameter extraction • 3-14
 using delay • 3-16
 parameter transformation • 2-23, @-1
 simple noise model • 6-15
 transition frequency • 3-14
 two-port transfer measurement • 3-11
 validity of hybrid-pi model • 3-19
 blackbox network representation • 2-10, 7-8
 – See also multi-ports • 2-11
 block diagram
 compensation • 5-5
 superposition flow diagram • 4-7
 superposition model • 4-6
 Boltzmann constant • 8-12, 8-19
 Butterworth response • 5-10, @-13

C

calibration
 of noise • 8-12
 shot noise calibrator • 8-17
 synthetic noise calibration • 8-13, 8-15
 using noise-tees
 at arbitrary reference plane • 8-21, 8-23
 at diode reference plane • 8-13, 8-15
 mathematical halving • 8-23
 of transfer measurements

PCB test-fixture one-ports • 3-6
 PCB test-fixture two-ports • 3-11
 cascading
 of noisy two-ports • 7-19
 two-ports • 2-19
 waves • 2-8
 cascading – See active decoupling • 4-25
 Chebyshev response • 5-10, @-13
 circuit stripping • 6-9
 closed loop gain – See loop gain • 5-3
 coaxial structure • 2-4
 coincident poles and zeros – See poles and zeros • 4-31, 5-23
 cold noise – See noise level • 8-5
 comb lines – See synthetic noise • 9-7
 compensation
 – See also aperture • 5-5, 5-10
 – See also superposition (synthesis) parameters • 5-5
 compensation algorithm • 5-25, @-16
 dominant compensation • 5-27, 5-33
 examples • 5-29, 5-33
 formal description • 5-5
 full compensation • 5-26, 5-29
 passive compensation • 5-6
 passive techniques • 5-13
 – See also pole shifting • 5-16
 – See also pole splitting • 5-17
 examples • 5-17, 5-18, 5-20
 phantom compensation • 5-6, 5-18
 profiled compensation • 5-6, 5-14
 intertwined solution • 5-16
 uncompensated loop gain • 5-6
 complex noise spectra – See spectra • 7-4
 consumed power – See power • 2-6
 continuous spectra – See spectra • 7-4
 conversion (input–,output–) – See superposition parameters • 4-5
 convolution • @-18
 correlation
 frequency domain
 – See correlation matrix • 7-10
 – See spectra of multiple signals • 7-6, @-18
 time domain • @-18
 – See auto correlation • 7-4
 correlation matrix
 composition • 7-11, 7-13
 history • 7-11, 7-13, 7-15
 of multi-ports • 7-6
 reduction of dimension • 7-12
 transformation • 7-11
 of thermal noise • 7-14
 of two-ports • 7-19
 cascading • 7-19
 transformation • 7-19
 relation with autonomous noise parameters • 7-25
 relation with spot noise parameters • 7-22

cross intensity spectrum – See spectra of multiple signals • 7-6
 cross-correlation – See spectra of multiple signals • 7-6
 current gain – See BJT • 2-23
 current-current feedback – See lightwave receivers • 6-2
 curve fitting
 – See also algorithms • @-5, @-6, @-9
 – See also extraction • 3-4
 constrained curve fits • 3-4
 Gauss-Newton • 3-3, 3-5, 4-24
 Levenberg-Marquardt • 3-3, 3-5, 4-24
 Padé approximation • 3-3
 polynomial fits • 3-3, 3-4, @-5
 rational delay fits • @-9
 rational function fits • 3-3, 3-5, 4-26, 4-32, @-6, @-9
 rational magnitude fits • 3-4, @-8
 rational phase fits • 3-4
 Remez algorithm • 3-3
 residual error term • 3-3

D

de-embedding • 2-19, 4-11
 deflation – See loop gain deflation • 4-29
 delay • 3-2
 in current gain of BJT • 3-16
 in transconductance of FET • 3-16
 pseudo delay
 – See extraction • @-9
 – See poles and zeros • 4-30, @-9
 desired system gain – See superposition parameters • 5-4
 device noise – See BJT and FET • 6-15
 discrete spectra – See spectra • 7-4
 distributed elements • 2-10
 division
 matrix division • @-2
 polynomial division • 4-31, @-10
 dominant compensation – See compensation • 5-27
 dominant singularities – See parasitic effects • 4-32
 dominant system order • 4-29
 double sided intensity spectra – See spectra • 7-3

E

effective aperture – See aperture • 5-3, 5-4
 Elementary charge • 8-12
 embedding • 2-19
 emitter inductance – See BJT • 3-14, 3-18
 ENR (excess noise ratio) – See noise level • 8-5
 equivalent input noise • 6-15
 – See also measurement (of noise) • 8-27
 – See also noise level • 7-7, 8-27
 calculation method • 7-7
 extracted from noise parameters • 7-20, 7-21, 8-49
 extraction algorithm • 8-29, @-22

- of electrical devices • 8-45
- of lightwave receivers • 8-32, 8-36
- excess noise
 - See noise level • 8-5
 - ...figure – See noise parameters • 7-22
 - ...ratio – See noise level • 8-5
- extraction
 - See also algorithms • 3-2
 - See also curve fitting • 3-2
 - See also de-embedding • 3-2
 - of input noise level • 8-29, @-22
 - of lumped element values • 3-4, 3-7, 3-14
 - of noise parameters • 8-47, @-24, @-27
 - of one-port parameters • 3-2
 - of poles and zeros • 3-5, 4-23
 - of superposition parameters • 4-11, 4-14
 - of transistor parameters • 3-14, 3-16
 - of two-port parameters • 3-10
 - pseudo delay • @-9
 - Taylor series expansion • 3-3, 3-14, 3-16, 3-18

F

- feedback analysis
 - See available bandwidth • 5-6
 - See feedback model • 4-3
 - See superposition parameters • 4-3, 5-6
- history • 1-2
- feedback factor – See superposition parameters • 4-5
- feedback model
 - asymptotic gain model • 4-3
 - elementary model of Black • 4-3
 - superposition model • 4-3
- feedback network – See superposition model • 4-4
- feedback synthesis
 - See aperture • 5-3
 - See compensation • 5-13, 5-22
 - See superposition parameters • 5-3
- iterative synthesis • 6-8, 6-14
- FET
 - parameter definition • 2-23
 - parameter extraction • 3-15
 - using delay • 3-16
 - parameter transformation • 2-23, @-1
 - simple noise model • 6-15
 - transconductance • 2-23, 3-15, 3-16
 - transition frequency • 3-15
 - two-port transfer measurement • 3-13
- forward amplifier – See superposition model • 4-4
- forward gain – See superposition parameters • 4-5
- forward leakage – See superposition parameters • 4-5
- Fourier spectra – See spectra • 7-4
- frame lines – See synthetic noise • 9-7
- front-end – See lightwave receivers • 6-2
- full compensation – See compensation • 5-26

G

- g-parameters • 2-18

- Gaussian distribution of noise • 9-24
- GB-product – See available bandwidth • 5-6
- generic passband – See aperture • 5-10, 5-23

H

- h-parameters • 2-18, 4-19, 7-18, @-1
- halving – See noise-tee • 8-23
- high frequencies • 1-5
- high impedance receivers – See lightwave receivers • 6-2
- higher order
 - ...approximation • 3-3
 - ...feedback loops • 1-7, 5-25, 5-29, 5-33, 6-7
 - ...mode excitation • 2-3, 2-4
 - ...polynomial fit • 3-3
 - ...rational function fit • 3-3
- dominant system order • 4-29
- history
 - of circuit simulators • 1-4
 - of feedback design • 1-2
 - of matrix noise parameters • 7-15
 - of matrix parameters • 2-11
 - of microwave electronics • 1-3
 - of noise correlation matrix • 7-11, 7-13, 7-15
 - of noise figure • 7-22
 - of noise measurements
 - lightwave receivers • 8-36, 9-2
 - noise parameters • 8-39
 - of noise parameters • 7-17, 7-21, 7-22
 - of synthetic noise generation • 9-2
 - of transfer measurements • 2-11, 3-2
- homodyne spectrum – See synthetic noise • 9-7
- hot noise – See noise level • 8-5

I

- impedance
 - See also z-parameters • 2-13
 - characteristic impedance • 2-5
 - impedance matrix • 2-13
 - impedance tuner – See measurements (of noise) • 8-10
 - load impedance • 2-6
 - reference impedance • 2-5
 - source impedance • 2-7
- incoherence threshold – See synthetic noise theory • 9-29
- injection of noise – See synthetic noise generation • 9-14
- input conversion – See superposition parameters • 4-5
- intensity spectra – See spectra • 7-4
- interconnection transformation
 - of two-ports • 2-19
 - of waves • 2-8
- intertwined profiled compensation – See compensation • 5-16
- iteration
 - See algorithms • 3-2
 - See curve fitting • 3-2

K

k-parameters • 2-18, 7-18

L

leakage (forward-) – See superposition parameters • 4-5

least squares sense • @-3

LHP (left half plane) • 4-23, 4-27, 4-30, 5-23

lightwave receivers

current-current feedback • 6-2, 6-4, 6-6

bandwidth performance • 6-13

noise performance • 6-21

high impedance receivers • 6-2, 6-3

integrating receivers • 6-3

low-impedance receivers • 6-2, 6-3

transimpedance receivers • 4-25, 6-2, 6-3

tuned receivers • 6-16

load model • 2-6

local feedback • 1-7, 5-17

long tailed pair – See active decoupling • 4-25

loop gain • 4-12, 5-6

– See also compensation • 5-13

– See also superposition parameters • 4-7

closed loop gain • 5-3

deflation • 4-29

deflation algorithm • @-11

pole-zero cancelation • 4-31

using pseudo delay • 4-29

one-cut method • 4-16, 4-17

approximation • 4-20

open loop gain • 4-16, 4-17

open loop current gain • 4-20

open loop voltage gain • 4-20

pole-zero extraction • 4-27

pole-zero pattern • 4-23, 4-27

profiled loop gain • 5-6

total loop gain • 5-6

two-cut method • 4-11, 4-16

loss-free conditions • 2-14

low-impedance receivers – See lightwave receivers • 6-2

LP-product – See available bandwidth • 5-6

lumped element models • 3-4

lumped elements • 2-10

M

mathematical halving

– See calibration (of noise) • 8-23

– See noise-tee • 8-23

matrix

– See also correlation matrix • 7-6, 7-10

– See also multi-ports • 2-11

– See also noise parameters • 7-8, 7-17

– See also two-ports • 2-18

– See also y, z, s, a, t-parameters • 2-11

multi-port transfer parameters • 2-11

port definition matrix • 2-15, 7-12, 7-13

wave interconnection matrix • 2-8

matrix algebra

matrix division • 2-13, @-2

transpose matrix • 2-12

measurement of noise

– See also calibration • 8-12

direct measurements • 8-3

extraction of input noise • @-22

extraction of noise parameters • @-24, @-27

for specific source impedance

arbitrary sources • 8-49

impedance tuner configuration • 8-10, 8-21

matched sources • 8-27, 8-29

mis-matched sources • 8-39, 8-45

full two-port noise measurements • 8-39

noise parameters • 8-47

state-of-the-art methods • 8-39

history • 8-39

improvements • 8-43

in lightwave receivers

alternative methods • 8-36

history • 8-36, 9-2

using synthetic noise • 8-32

pitfalls and accuracy • 8-3, 8-4, 9-24, 9-25

ratio measurements • 8-3, 8-27, 8-32

using spectrum analyzers • 7-6, 8-3, 8-4, 9-24, 9-25

measurement of transfer

calibration • 3-6, 3-11

history • 2-11

one-ports • 3-6

two-ports • 3-11

microstrip structure • 2-4, 2-11, 3-6, 3-11, 8-6

minimum noise figure – See noise parameters • 7-22

minimum phase response • 5-24, @-8

mirrored poles and zeros – See poles and zeros • 4-27, 5-23

multi mode signal flow • 2-2

multi-port noise analysis

– See also correlation matrix • 7-10

– See also noise parameters • 7-10

blackbox description method • 7-8

history of noise parameters • 7-14, 7-17, 7-21, 7-22

matrix transformation rules • 7-10

thermal noise • 7-14

multi-port transfer analysis

blackbox description method • 2-10

history of matrix parameters • 2-11

matrix properties • 2-13

matrix transfer parameters • 2-11

matrix transformation rules • 2-12

port reduction algorithm • 2-15

multi-ports

– See also measurement • 3-11

– See also two-ports • 2-18

N

NCR (noise current ratio) – See noise level • 8-13
 noise analysis – See spectral noise analysis • 7-7
 noise correlation matrix – See correlation matrix • 7-6
 noise current ratio (NCR) – See noise level • 8-13
 noise figure – See noise parameters • 7-22
 noise injection – See synthetic noise generation • 9-14
 noise level
 – See also (intensity) spectra • 7-4
 – See also calibration • 8-12
 – See also equivalent input noise • 7-7
 – See also measurement • 8-27
 – See also noise parameters • 8-5
 – See also thermal noise • 8-5
 related with calibrated sources
 excess noise ratio (ENR) • 8-5, 8-29
 noise current ratio (NCR) • 8-13
 shot noise level • 8-12
 standard reference temperature • 8-5
 thermal noise level • 8-12
 related with source temperature
 cold noise • 8-5, 8-8
 excess noise • 8-5, 8-8
 hot noise • 8-5, 8-8
 relative intensity noise (laser RIN) • 8-13
 noise measure – See noise parameters • 7-24
 noise model – See BJT and FET • 6-15
 noise optimization • 6-15
 – See also (spot) noise parameters • 7-22
 noise tuning • 6-16
 noise tuning with feedback • 6-20
 noise parameters
 – See also correlation matrix • 7-10
 – See also noise level • 8-5
 autonomous noise parameters • 7-24
 extraction algorithm • @-24, @-27
 history • 7-21
 matrix noise parameters • 7-8, 7-17
 history • 7-15
 measurement methods • 8-39, 8-47
 spot noise parameters • 7-21
 excess noise figure • 7-22, 8-39
 history • 8-39
 minimum noise figure • 7-22
 noise figure • 7-22, 8-39
 noise measure • 7-24
 optimal source impedance • 7-22
 thermal noise • 7-14
 transformation rules • 7-22
 noise source
 – See also noise level • 8-5
 – See also noise-tee • 8-6
 – See also shot noise • 8-15
 – See also synthetic noise • 9-1
 – See also thermal noise • 8-15
 hot/cold 50 ohm source
 application • 8-13
 noise extraction algorithm • 8-29
 performance • 8-9

multi-level noise source • 8-6
 matched configuration • 8-7
 mis-matched configuration • 8-10, 8-21, 8-45
 noise extraction algorithm • 8-29
 performance • 8-9
 noise spectra – See spectra • 7-4, 7-6
 noise temperature – See thermal noise • 8-5
 noise tuning – See noise optimization • 6-16, 6-20
 noise-tee • 8-6
 – See also calibration • 8-13
 matched configuration • 8-7, 8-13
 output impedance performance • 8-9
 output level performance • 8-8
 mathematical halving • 8-23
 mismatched configuration • 8-10, 8-21, @-22
 normal distribution of noise • 9-24
 Norton source model • 2-7, 7-9

O

one-cut method – See loop gain • 4-16
 one-ports
 load model • 2-6
 source model • 2-7
 open loop gain – See loop gain • 4-16
 output conversion – See superposition parameters • 4-5
 overall feedback • 1-7, 6-7, 6-14, 6-20

P

parameters
 – See multi-ports • 2-11
 – See noise parameters • 7-8
 – See superposition parameters • 4-5
 – See two-ports • 2-18
 parasitic effects • 1-5, 2-1, 4-1
 – See also active decoupling • 5-1
 dominant singularities • 4-32, 5-9
 in intertwined profiled compensators • 5-16, 5-17
 in phantom compensators • 5-19
 parasitic singularities • 4-29, 5-33
 quantification of parasitics • 6-9, 6-12
 Parseval theorem • 7-5, @-20
 passband analysis/synthesis
 – See available bandwidth • 5-9
 passband analysis/synthesis – See aperture • 5-9
 phantom compensation – See compensation • 5-5, 5-6, 5-18
 phantom zero – See zeros, compensation • 5-5
 Planck constant • 8-19
 pole shifting • 5-15, 5-16
 pole splitting • 5-15, 5-17
 pole-zero pattern – See poles and zeros • 4-23
 poles and zeros
 – See also loop gain deflation • 4-29
 coincident poles and zeros • 4-31, 5-23
 definition • 3-2
 dominant singularities • 4-32, 5-9

extraction algorithm • 3-5, @-6, @-8, @-9
 extraction example • 4-26
 extraction strategy • 4-23
 LHP-poles and zeros • 4-30, 5-23
 mirrored poles and zeros • 4-27, 5-23
 of phantom compensator • 5-18
 of profiled compensator • 5-15
 in intertwined solutions • 5-16
 parasitic singularities • 4-29
 pole-zero pattern • 4-27
 interpretation • 4-27
 of all-pass response • 5-23
 of Bessel response • 5-10
 of Butterworth response • 5-10
 of Chebyshev response • 5-10
 of compensated loops • 5-29, 5-33
 of generic passband • 5-23
 of low-pass response • 5-10, 5-23
 of pseudo delay response • 4-30, 5-23
 primary singularities • 4-29
 pseudo delay • @-9
 quantification of parasitics • 4-25, 6-12
 root-locus • 5-29, 5-33
 polishing of polynomials • 4-31
 polynomial division • @-10
 polynomial fits – See curve fitting • 3-4, @-5
 port definition matrix – See matrix • 2-15, 7-13
 power
 average power flow • 2-5
 complex power waves • 2-5
 consumed load power
 in multi-ports • 2-12
 in one-ports • 2-6
 exchangeable source power
 of multi-ports • 7-9
 of one-ports • 2-8
 power spectra – See (intensity) spectra • 7-4
 primary singularities – See poles and zeros • 4-29
 profiled compensation – See compensation • 5-6, 5-14
 profiled loop gain – See loop gain • 5-6
 pseudo delay
 – See extraction • @-9
 – See poles and zeros • 4-30
 PZ-pattern – See poles and zeros • 4-23

R

rational functions
 curve fitting • 3-5, @-6, @-8, @-9
 definition • 3-2
 realized system gain – See superposition parameters • 4-7
 receivers – See lightwave receivers • 6-2
 reciprocal conditions • 2-13
 reference impedance
 definition • 2-5
 loads • 2-6
 multi-ports • 2-13
 sources • 2-7

reference plane
 interconnection matrix • 2-8
 on PCB • 2-4, 2-11, 3-6, 3-11, 6-10
 orientation • 2-6, 2-7, 2-8
 reference temperature – See thermal noise • 8-5
 reflection
 load reflection • 2-6
 reflection tuner – See measurements (of noise) • 8-10
 source reflection • 2-7
 relative intensity noise (RIN) – See noise level • 8-13
 resolution bandwidth • 7-5
 RHP (right half plane) • 4-23, 4-27, 4-30, 5-23
 RIN (relative intensity noise) – See noise level • 8-13
 room temperature – See thermal noise • 8-5
 root-locus – See poles and zeros • 5-29, 5-33

S

s-parameters • 2-13, 2-14, 2-18, 4-19, 7-9, 7-14, 7-18, @-1
 scattering
 – See also s-parameters • 2-13
 load reflection • 2-6
 scattering matrix • 2-13
 source reflection • 2-7
 self intensity spectrum – See spectra of multiple signals • 7-6
 self-heterodyning – See synthetic noise generation • 9-17
 self-homodyning – See synthetic noise generation • 9-17
 shot noise • 8-12, 8-15
 shot noise generation
 illumination limit • 8-18
 using incandescent lamps • 8-16
 using laser illumination • 8-19
 using LED illumination • 8-19
 signal flow • 2-3
 – See also power • 2-5
 parameters • 2-5
 reference impedance • 2-5
 reference plane • 2-5
 single and multi mode • 2-2
 voltage and current • 2-5
 wave definition • 2-5
 single mode signal flow • 2-2, 2-11
 single sided intensity spectra – See spectra • 7-3
 single-cut method – See loop gain • 4-16
 singularities – See poles and zeros • 4-29
 source model
 – See also noise source • 8-6
 of multi-ports • 7-9
 of one-ports • 2-7
 thermal noise • 7-21
 specific passband – See aperture • 5-23
 spectra
 – See also Parseval theorem • 7-5

- See also synthetic noise • 9-4
 - See also Wiener-Khintchine theorem • 7-5
 - complex noise spectra $N(f)$ • 7-4, @-18
 - continuous spectra $F(f)$ • 7-4, @-18
 - conventions • 7-3, @-18
 - definitions • 7-4, @-18
 - discrete spectra $Q(n)$ • 7-4, @-18
 - intensity spectra • 7-4, @-18
 - double sided $D(f)$ • 7-3, 7-5
 - single sided $S(f)$ • 7-3, 7-5
 - spectra of multiple signals
 - See also correlation matrix • 7-6
 - correlation coefficient • 7-6
 - cross correlation • 7-6
 - cross intensity spectrum • 7-6
 - self intensity spectrum • 7-6
 - Wiener-Khintchine theorem in matrix notation • 7-7
 - spectral noise analysis
 - automated methods • 7-13
 - manual methods • 7-7
 - spectrum analyzer
 - See also measurement of noise • 8-3
 - as noise detector • 8-3, 8-4
 - noise accuracy limits • 8-4, 9-24, 9-25
 - readout • 7-6
 - resolution bandwidth • 7-6
 - video bandwidth • 7-6
 - spot noise parameters – See noise parameters • 7-21
 - standard reference temperature – See thermal noise • 8-5
 - Stefan-Boltzmann constant • 8-19
 - stripline structure • 2-4, 8-6
 - superposition model
 - definition • 4-4
 - example • 4-9
 - feedback network • 4-4
 - flow diagram • 4-7
 - forward amplifier • 4-4
 - signal flow • 4-3
 - superposition parameters
 - flow parameters
 - See also loop gain • 4-7
 - asymptotic system gain • 4-7, 4-14, 5-4
 - extraction • 4-9, 4-11, 4-14
 - loop gain • 4-7, 4-12
 - realized system gain • 4-7, 5-4
 - variable system gain • 4-14
 - virtual system gain • 4-7, 4-14, 5-4
 - virtual-aperture • 5-4
 - model parameters
 - feedback factor • 4-5, 4-14
 - forward gain • 4-5, 4-13
 - forward leakage • 4-5, 4-7, 4-14, 5-4
 - input conversion • 4-5
 - loop gain • 5-6
 - output conversion • 4-5
 - synthesis parameters
 - See also aperture • 5-4
 - desired system gain • 5-4
 - effective-aperture • 5-4
 - phantom compensation • 5-4
 - profiled compensation • 5-4
 - profiled loop gain • 5-6
 - total loop gain • 5-6
 - synthetic noise
 - See also calibration • 8-12
 - See also noise level • 8-12
 - definitions
 - comb lines • 9-7
 - frame lines • 9-7
 - homodyne spectrum • 9-7, 9-17
 - self-heterodyning • 9-17
 - self-homodyning • 9-17
 - synthetic noise generator • 9-7
 - synthetic noise spectrum • 9-7
 - electrical applications • 8-7, 8-9, 8-13, 8-15, 8-21
 - history • 9-2
 - lightwave applications • 8-32
 - noise current ratio (NCR)
 - approximation • 8-12, 9-20
 - definition • 8-12
 - synthetic noise generation
 - basic principle • 9-4
 - power analysis • 9-20
 - dc-current • 9-20
 - rms-magnitude • 9-20
 - spectral intensity • 9-20
 - spectral analysis • 9-11
 - noise injection • 9-14
 - output spectrum • 9-7, 9-11, 9-18
 - simulated frame lines • 9-11
 - stabilization • 9-22
 - statistical analysis • 9-22
 - synthetic noise theory
 - incoherence analysis • 9-29
 - incoherence condition • 9-32
 - incoherence sensitivity • 9-31
 - incoherence threshold • 9-29
 - power analysis • 9-30
 - spectral analysis • 9-34
 - statistical analysis • 9-38
 - time domain analysis • 9-26
 - system gain – See superposition parameters • 4-7
- ## T
- t-parameters • 2-18, 4-19, 7-18, @-1
 - Taylor series expansion – See extraction • 3-14
 - temperature – See thermal noise • 8-5
 - thermal noise • 7-14, 8-5
 - See also noise figure • 8-5
 - See also noise level • 8-5
 - for measurement purposes • 8-8, 8-12, 8-36
 - in multi-ports – See correlation matrix • 7-14
 - in resistors • 8-12
 - in sources • 7-21, 7-22, 8-22
 - standard reference temperature • 8-5

Thévenin source model • 2-7, 7-9
 total loop gain – See loop gain • 5-6
 transconductance – See FET • 2-23
 transfer
 – See extraction • 3-2
 – See multi-ports • 2-11
 – See poles and zeros • 3-2
 – See two-ports • 2-18
 transimpedance receivers – See lightwave receivers
 • 6-2
 transistors – See BJT or FET • 2-22
 transition frequency • 4-27, 5-2
 – See also BJT and FET • 3-14, 4-1
 transmission line structures • 2-4
 trigonometric distribution of noise • 9-24
 tuned receivers
 – See lightwave receivers • 6-16
 – See noise optimization • 6-16
 two-cut method – See loop gain • 4-16
 two-port noise analysis
 – See correlation matrix • 7-21
 – See noise parameters • 7-21
 two-port transfer analysis
 matrix parameters • 2-18
 matrix transformation rules • 2-19, @-1
 virtual circuit parameters • 2-20
 two-ports
 – See also measurement • 3-11
 – See also multi-ports • 2-11

V

virtual aperture – See aperture • 5-3
 virtual circuit model • 2-21, 3-14, 3-15, 4-9, 4-13,
 4-15, 4-18
 virtual circuit parameters

general concept • 2-20
 of feedback loops • 4-13
 of photo diodes • 3-5
 of transistors • 2-22, 3-11, 3-13
 virtual system gain – See superposition parameters
 • 4-7
 virtual-aperture – See aperture • 5-4

W

waves
 definition • 2-5
 interconnection matrix • 2-8
 wave model of loads • 2-6
 wave model of noise • 7-9, 7-20, 7-21
 wave model of sources • 2-7, 7-9, 7-18
 wideband • 1-5
 Wiener-Khintchine theorem • 7-5, @-21
 in matrix notation • 7-7

Y

y-parameters • 2-13, 2-14, 2-15, 2-16, 2-18, 2-21,
 3-14, 3-15, 4-10, 4-14, 4-18, 4-19, 4-21, 7-9,
 7-12, 7-13, 7-14, 7-18, @-1

Z

z-parameters • 2-13, 2-14, 2-15, 2-16, 2-18, 2-21,
 7-9, 7-12, 7-13, 7-14, 7-18, @-1
 zeros
 compensating zeros • 5-15, 5-16
 phantom zeros • 5-5, 5-7, 5-18, 5-29, 5-33
 profiled zeros • 5-15
 RHP-zeros • 4-27, 4-30, 5-23